Original Paper

Effect of AI-Based Natural Language Feedback on Engagement and Clinical Outcomes in Fully Self-Guided Internet-Based Cognitive Behavioral Therapy for Depression: 3-Arm Randomized Controlled Trial

Mirai So[1*], MD, MBA, PhD; Yoichi Sekizawa[2*], MA; Sora Hashimoto[3*], MA; Masami Kashimura[4*], PhD; Hajime Yamakage[5*], MEng; Norio Watanabe[6*], MD, PhD

[1]Department of Psychiatry, Tokyo Dental College, Tokyo, Japan

[2]Research Institute of Economy, Trade and Industry, Tokyo, Japan

[3]United Health Communication Co., Ltd., Tokyo, Japan

[4]Department of Psychology, Faculty of Human Schiences, Tokiwa University, Ibaraki, Japan

[5]Department of Medical Statistics, Satista Co., Ltd., Kyoto, Japan

[6]Department of Psychiatry, Soseikai General Hospital, Kyoto, Japan

[*]all authors contributed equally

Corresponding Author:
Mirai So, MD, MBA, PhD
Department of Psychiatry
Tokyo Dental College
2-9-18 Misakicho, Chiyoda-ku
Tokyo, 101-0061
Japan
Phone: 81 47 322 0151
Fax: 81 47 325 4456
Email: mirai.so@keio.jp

Abstract

Background: Depression remains a major global cause of disability; yet, access to optimal mental health services is limited. Self-guided internet-based cognitive behavioral therapy (iCBT) offers a scalable alternative but is generally less effective than guided programs, showing limited antidepressant effects and incomplete symptomatic and functional recovery. Adherence remains a major barrier. Recent advances in artificial intelligence (AI), particularly natural language processing, enable automated advisory and empathic feedback that may enhance engagement and therapeutic impact. Although previous trials have reported promising effects, most used heterogeneous control conditions, making it difficult to isolate the specific contribution of AI within fully self-guided interventions.

Objective: This randomized controlled trial evaluated whether natural language processing–based AI feedback integrated into a fully self-guided iCBT program improves clinical outcomes and engagement compared with an otherwise identical iCBT program without AI support.

Methods: We recruited 1187 adults aged 20-60 years online and randomly assigned them to AI-augmented iCBT (AI-iCBT; n=396), iCBT without AI (n=397), or a waitlist control (n=394). Both active groups received 6 weekly sessions combining video-based psychoeducation and cognitive restructuring exercises. The AI-iCBT program additionally provided automated empathic and advisory feedback. The primary outcome was depressive symptom severity (Patient Health Questionnaire-9 [PHQ-9]) at week 7 and month 3, analyzed using mixed-effects models for repeated measures under an intention-to-treat framework. Secondary outcomes included a dichotomous PHQ-9 score of ≥10, Quick Inventory of Depressive Symptomatology, Generalized Anxiety Disorder-7, Sheehan Disability Scale, and weekly participation rates. Exploratory analyses assessed the impact of AI functions on engagement and antidepressant effects in the efficacy analysis set (EAS).

Results: In intention-to-treat analyses, no significant between-group differences were observed in mean PHQ-9 scores at week 7 or month 3, whereas engagement analyses showed a significant group × week interaction, with AI-iCBT participants demonstrating

consistently higher odds of weekly participation (odds ratio 1.23, 95% CI 1.09-1.39; *P*<.001). Exploratory analyses indicated that activation of the empathic feedback function strongly predicted adherence (odds ratio 9.99, 95% CI 5.80-17.21; *P*<.001), while advisory feedback was not significant. In EAS analyses, iCBT showed significant short-term improvement versus control at postintervention, whereas at follow-up, only AI-iCBT showed a significantly lower proportion of participants with a PHQ-9 score of ≥10 compared with control (difference –0.15, 95% CI –0.30 to –0.01; *P*=.046). No serious adverse events were reported.

Conclusions: AI support significantly improved adherence to a fully self-administered program. In EAS analyses, AI-iCBT also showed a significantly lower proportion of participants with PHQ-9 score of ≥10 at follow-up compared with control. Empathic feedback emerged as a key mechanism for sustaining engagement, suggesting that AI communication may help maintain participation in scalable digital mental health interventions. Further research is required.

Introduction

Depression is a leading global cause of disability [1], substantially impairing quality of life [2,3] and imposing a considerable economic burden, including medical expenses [4] and productivity losses [5]. The rising prevalence of depression, combined with a shortage of health care resources, places a significant strain on health systems and professionals in meeting the growing demand [6-9].

In response, technology-delivered self-help interventions have emerged as promising solutions for managing mental health difficulties. Most of these interventions are based on cognitive behavioral therapy (CBT) and are generally referred to as internet-based CBT (iCBT) [10,11]. iCBT can be delivered either with therapist support (guided) or without therapist support (unguided, self-directed). Compared to traditional face-to-face therapy, iCBT offers major advantages in terms of accessibility, availability, and cost-effectiveness for both patients and providers. Furthermore, its online format provides benefits related to privacy, confidentiality, and anonymity, which can help reduce the stigma often associated with seeking mental health care [12,13].

While iCBT produces clinically meaningful symptom improvements, remission rates tend to be modest (approximately 30%-35%). An individual participant data meta-analysis reported a remission rate of 35.2% and a response rate of 56% [14]. Large-scale individual participant data network meta-analyses have consistently shown that guided iCBT yields higher response and remission rates than unguided formats, reflecting the challenges of engagement and dropout in fully self-administered programs [15-19].

This study specifically examines the unguided, fully self-administered format. Such interventions enable users to manage their symptoms independently and offer potential benefits such as reducing costs, alleviating the burden on health care providers, and improving access to mental health services in regions where such services are difficult to obtain [9,16]. Previous research has demonstrated that sociodemographic factors, such as age and sex, are associated with dropout risk in iCBT [20-22].

Nevertheless, if the effectiveness and engagement of unguided iCBT could be enhanced, the benefits of structured self-help materials would not be limited to fully self-administered interventions but would also extend to guided and blended formats. When patients acquire skills and knowledge through self-help modules, they can participate more effectively in therapist-led sessions, thereby enhancing overall treatment outcomes [23]. Strengthening such "self-help effects" not only amplifies therapeutic gains in guided and blended care but also reduces the time and workload required from therapists. By improving scalability and cost-effectiveness, it further increases the feasibility of implementation in routine practice [24,25]

To address these challenges, natural language processing (NLP), an artificial intelligence (AI) technology that enables the understanding and generation of human language, has increasingly been applied to enhance adherence and engagement through tailored feedback [26-28]. Whereas conventional unguided iCBT typically provides static or generic responses, in this study, we used an NLP-enabled iCBT program with automated advisory and empathic functions. This allows the system to generate advisory and empathic feedback in response to user input, potentially addressing both emotional and procedural barriers simultaneously.

Despite its promise, the specific therapeutic contributions of NLP remain unclear. Previous studies have frequently used heterogeneous control conditions such as waitlists, no intervention, treatment as usual, bibliotherapy, or conversational computer programs [28-32]. This heterogeneity makes it difficult to determine whether NLP provides distinct therapeutic benefits or merely functions as an active placebo by enhancing user expectations.

Against this background, the aim of this study was to conduct a randomized, parallel-group exploratory trial directly comparing 2 unguided, fully self-administered iCBT programs that were identical except for the presence or absence of NLP feedback. This design allowed us to evaluate the therapeutic contribution

of NLP within a self-help framework in a blinded comparison of the 2 intervention groups.

Methods

Overview

This study was a 3-arm randomized controlled trial, with double-blinding between the AI-augmented iCBT (AI-iCBT) and iCBT groups, while the waitlist control group was unblinded. The intervention arms consisted of AI-iCBT, an unguided, fully self-administered iCBT program incorporating NLP feedback, and iCBT, an unguided, fully self-administered program without NLP feedback. These 2 arms are hereafter collectively referred to as "unguided iCBT." The waitlist group served as the control condition.

Study Participants

Invitation emails were sent to all monitors registered with Nikkei Research Inc, with the aim of recruiting at least 900 participants. Interested individuals were directed to complete an online screening survey, which included the Patient Health Questionnaire-9 (PHQ-9) [33,34], to determine eligibility.

Eligibility Criteria

Inclusion and exclusion criteria are provided in Textbox 1.

Textbox 1. Eligibility criteria.

Inclusion criteria

- Aged 20-60 years (to target the working-age population and exclude older adults with lower digital literacy)
- Had access to the internet
- Baseline Patient Health Questionnaire-9 score of ≥5 (this threshold was selected to avoid floor effects and ensure adequate symptom levels for change detection) [33,35,36]
- Ability to understand Japanese

Exclusion criteria

- Presence of medical conditions precluding participation as determined by a physician
- Concurrent participation in another cognitive behavioral therapy program
- Diagnosis of schizophrenia
- Severe suicidal ideation
- Diagnosis of dementia
- Substance dependence in the past 12 months (excluding smoking)

Eligible participants were randomly assigned to 1 of the 3 groups. Immediately before the intervention, PHQ-9 scores were reassessed; individuals with scores <5 were excluded from the efficacy analysis set (EAS) but remained in the overall study population. The detailed definition of the EAS is provided in the Analytic Strategy section.

Intervention

The AI-augmented iCBT program, developed by NEC Solution Innovators, Ltd (Tokyo), integrates an NLP module trained on 28,718 prior iCBT records from Japanese users. The entire program was delivered in Japanese, and Figure 1 presents an English-translated version of the original Japanese interface for publication purposes. The program includes a self-guided cognitive restructuring (CR) exercise, where users complete a 7-column thought record to address cognitive distortions. The NLP system processes user inputs, including situation, automatic thoughts, and feelings, referencing a corpus of past responses (Figure 1). It provides 2 types of automated feedback with text and phonation: (1) empathetic messages delivered through an animated character whose expressions, such as smiling or showing sadness, are synchronized with the message content, and (2) advisory messages offering guidance to refine inputs or direct users to appropriate exercises, including suggestions to

revise content if the user's input was unclear or misplaced (eg, a feeling given instead of a thought; Figure 2).

In contrast, the non-AI iCBT program retained the same structure but provided only neutral, noncontextual responses, such as generic phrases like "Uh-huh" with neutral facial expressions. Both AI-iCBT and iCBT programs were otherwise identical in content, using the same validated 6-session self-guided iCBT package. This iCBT program has previously demonstrated significant antidepressant effects in a randomized controlled trial among working adults (n=60 per group), compared with a waitlist control, with a medium to large effect size (Cohen d=0.65; P<.005) [37]. This package consisted of 6 weekly sessions, each including a 15-minute video-based psychoeducation module covering standard CBT principles such as behavioral activation and problem-solving, along with a weekly CR exercise in which users applied learned techniques. In this trial, the only difference was the addition of the NLP feedback system. The program was available on both smartphones and PCs. The AI-enhanced features, which were designed in advance to improve user engagement and response accuracy, exhibited high usability, with low dissatisfaction rates reported for both the empathetic (3/32, 9.4%) and advisory (1/24, 4.2%) feedback functions (Figure 3).

Figure 1. Examples of expressions extracted from the natural language processing corpus and categorized into 4 domains: Problem, Trouble, Feeling, and Subjective.
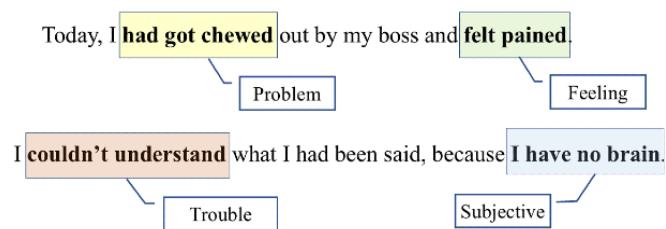

Figure 2. Workflow of artificial intelligence-guided internet-based cognitive behavioral therapy (CBT), showing the structured 7-step cognitive restructuring exercise with automated prompts and feedback. AT: automatic thought; NLP: natural language processing.
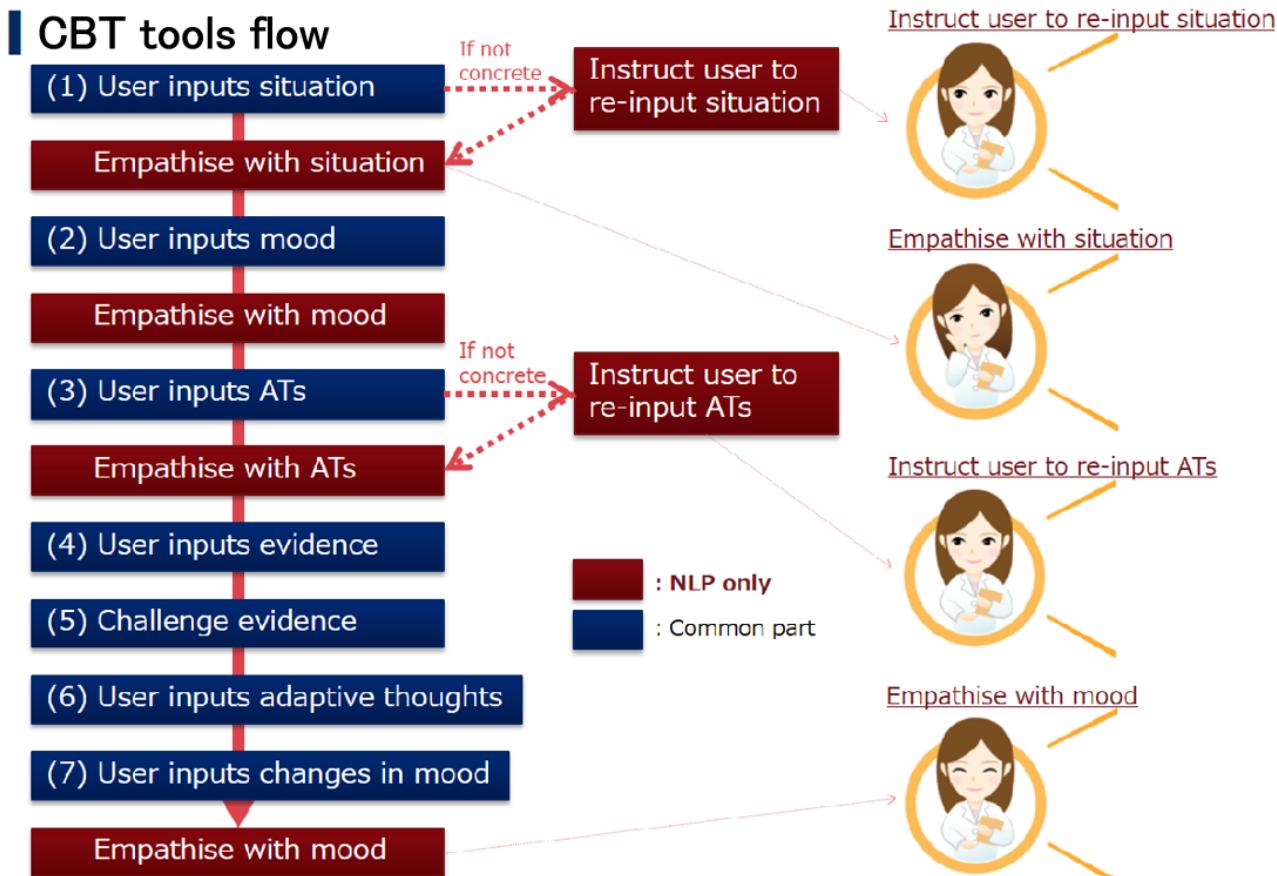
Figure 3. User acceptability ratings of natural language processing–generated feedback for empathy and automatic thought identification. AT: automatic thought.
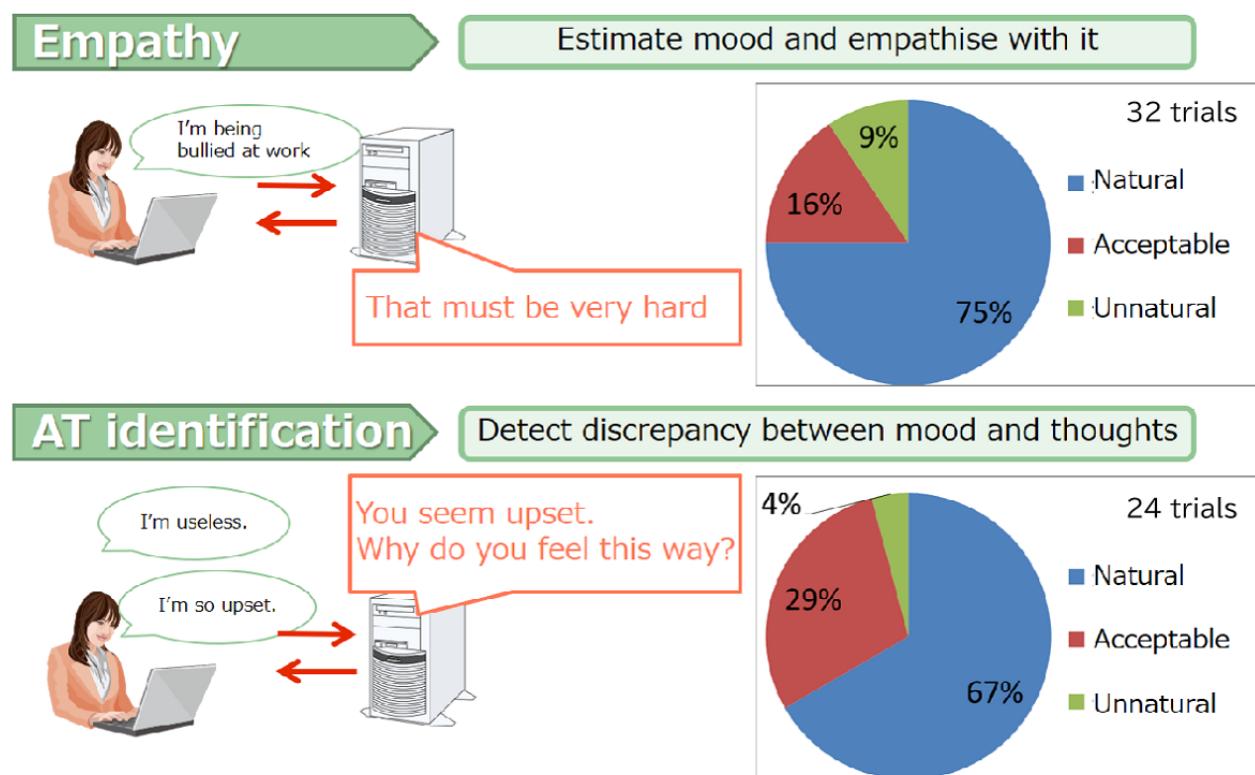
Randomization and Masking

The registered participants were randomly and concurrently assigned to either the AI-iCBT, iCBT, or waitlist groups using a computer-generated random sequence provided by an independent third party. Stratified randomization was applied based on age (≤40 vs >40 years), sex, and baseline PHQ-9 score (≤9 vs >9), as baseline symptom severity has been shown to influence treatment outcomes in self-guided iCBT [23]. Participants in the waitlist group were aware of their allocation and were therefore unblinded, whereas those in the AI-iCBT and iCBT groups were told only that they would participate in iCBT using "the latest technology," without disclosure of their specific group assignment. Accordingly, blinding was implemented between the 2 intervention groups.

Study Procedures

Automated email reminders were sent to participants twice weekly during the 7-week intervention period. Each week, participants in the intervention groups were required to (1) view an online psychoeducational CBT module and (2) perform their allocated (AI-iCBT or iCBT) CR exercise at least once (6 times or more in total). Waitlist participants did not undergo any exercises during this period. All participants were required to complete assessments at baseline, postintervention (week 7), and follow-up (month 3 after postintervention). All intervention and assessment procedures, including attendance and outcome measures, were conducted online.

Outcomes

All primary and secondary outcomes were analyzed based on the intention-to-treat (ITT) population, which included all randomized participants.

Primary Outcome

The primary outcome was the mean PHQ-9 score, assessed at baseline, week 7 (postintervention), and month 3 (follow-up). The PHQ-9 is a widely used self-report measure of depressive symptoms (range 0-27, higher scores indicating greater severity), originally developed by Kroenke et al [33] and validated in Japanese [38].

Secondary Outcomes

Secondary outcomes include (1) proportion of participants with PHQ-9 scores ≥10 (a conventional cutoff for probable major depression) [33,39]. Although not selected as the primary outcome in this study, such binary outcomes are often considered clinically meaningful, as they reflect remission from a diagnostic threshold [40-43]. (2) Quick Inventory of Depressive Symptomatology-Japanese version (QIDS-J) [43,44]—a self-report scale of depressive symptom severity. (3) Generalized Anxiety Disorder-7 (GAD-7) [45,46]—a self-report questionnaire measuring generalized anxiety symptoms. (4) Sheehan Disability Scale (SDS) [47,48], which evaluates functional impairment in work, social, and family life, with SDS ≥10 commonly used as a pragmatic threshold in clinical trials [49].

Engagement outcome included weekly CR exercise attendance rate (defined as attending at least one session per week) in the 2 intervention groups. Program satisfaction at week 7 was

assessed with the Client Satisfaction Questionnaire-8 (CSQ-8) [50,51], for which the Japanese version has demonstrated reliability and validity.

All outcomes except engagement and satisfaction were assessed at baseline, week 7, and month 3.

Analytic Strategy

Overview

In this study, all primary and secondary analyses were conducted in the ITT population, defined as all randomized participants. In addition, 2 exploratory analyses were performed: (1) as an ad hoc exploratory analysis, we examined which AI feedback function (empathy or advisory) contributed more to enhancing engagement, and (2) as an additional exploratory analysis, we assessed continuous and binary PHQ-9 outcomes within the EAS.

Engagement-Enhancing Factors

For this analysis, the 2 intervention groups were combined, and the presence or absence of empathy and advisory feedback during week 1 was examined in relation to engagement from weeks 2 to 6, defined as completing at least 1 exercise per week. The detailed statistical methods are described in the Statistical Analysis section.

Efficacy Analysis Set

The EAS was defined as participants with a baseline PHQ-9 score of ≥5 and completion of at least 3 out of the 6 weekly sessions. Participants with a baseline PHQ-9 score of <5 (minimal symptoms) were excluded, as their inclusion could reduce the power to detect change and dilute the mean effects [52,53]. Furthermore, previous research has demonstrated a dose-response relationship in iCBT, with clinical benefits emerging after completing approximately half of the modules; therefore, the minimum attendance criterion was set at 3 of 6 sessions [54,55].

Statistical Analysis

Overview

The sample size was estimated based on an assumed effect size of 0.10 (Cohen *d*) between the AI-iCBT and iCBT groups, given the absence of directly comparable prior studies. A dropout rate of 50% was anticipated based on patterns observed in similar previous studies. The power was set at 80% with a 2-sided significance level of $\alpha=.05$. As this was an exploratory study, no adjustment for multiplicity was applied, and nominal *P* values were reported.

The primary and secondary analyses were conducted according to the ITT principle, including all randomized participants. Baseline demographic and clinical characteristics were compared across groups using 1-way ANOVA or chi-square tests.

Continuous outcomes (PHQ-9, QIDS-J, GAD-7, and SDS) were analyzed using a mixed-effects model for repeated measures (MMRM), with intervention, time, and intervention × time interaction as fixed effects, assuming an unstructured covariance structure. Results are presented as least squares means with 95% CIs.

Binary outcomes (PHQ-9 ≥10) were analyzed using generalized linear mixed models (GLMMs) with a binomial distribution and logit link, including intervention, time, and their interaction as fixed effects, and subject as a random effect. Estimated proportions and their 95% CIs were reported. Missing data for the outcomes were handled under the missing at random assumption within the MMRM and GLMMs framework.

CR exercise participation rates (defined as at least 1 completion per week) in the intervention groups were analyzed using GLMMs with a logit link, including intervention, week (as a continuous variable), and intervention × week interaction as fixed effects.

Exploratory Analyses

The following two exploratory analyses were conducted.

Engagement-Enhancing Factors

The dependent variable was defined as achieving at least 1 CR exercise per week across all weeks from week 2 to week 6 (yes/no). Independent variables were the presence or absence of empathy or advisory feedback during week 1. Covariates included age, sex, marital status, education, employment status, history of psychiatric and physical treatment, baseline PHQ-9 score, and intervention group (AI-iCBT vs iCBT), as group differences could confound the association of interest. Analyses were performed using generalized estimating equations logistic regression models to account for within-subject correlation and to estimate population-averaged effects.

Efficacy Analysis Set

In the EAS (participants with a baseline PHQ-9 score of ≥5 and completion of ≥3 sessions), continuous and binary PHQ-9 outcomes were additionally adjusted for age, sex, baseline PHQ-9 score, and medical history as covariates to account for potential group imbalances in the restricted sample.

Sensitivity Analyses

Associate Factors of Low Adherence

To explore potential factors associated with dropout, we compared baseline characteristics between EAS participants who attended ≥3 sessions and those who attended <3 sessions, given the high attrition typically observed in self-guided digital interventions.

Alternative Definition of Caseness

We conducted an exploratory analysis on the binary PHQ-9 outcome in the EAS, applying a stricter definition of depression severity: a PHQ-9 score of ≥10 plus at least 1 core symptom (depressed mood or anhedonia) [56,57], together with an SDS score of ≥10 as an indicator of functional impairment [49,58-60].

All analyses were performed using IBM SPSS Statistics version 26.0.

Reporting Standards

Reporting of this trial followed the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement [61] and the

CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [62] for internet-based interventions. The completed CONSORT-EHEALTH checklist is submitted as Multimedia Appendix 1.

Ethical Considerations

This study was reviewed and approved by the Hiramatsu Memorial Hospital Ethics Committee (approval number 20150807). All participants provided informed consent electronically prior to enrollment after reading an online information sheet describing the study purpose, procedures, potential risks, and voluntary nature of participation. Participants were informed that they could withdraw at any time without penalty.

The trial was prospectively registered in the University Hospital Medical Information Network Clinical Trials Registry.

All data were anonymized prior to analysis to ensure confidentiality. No personally identifiable information was accessible to the research team. Participants who completed the final assessment received a ¥500 (US $4.5) gift voucher as compensation. No identifiable images or other personal data are presented in this manuscript.

Results

Study Participants

A total of 1187 participants were eligible and randomly allocated to the AI-iCBT (n=396), iCBT (n=397), or waitlist (n=394) groups (ITT population; see Figure 4 for the CONSORT flow diagram). Baseline demographic and clinical characteristics are summarized in Table 1. The mean age was 43.50 (SD 9.85) years, and 699 (58.8%) of 1187 participants were male. Across the 3 groups, demographic and clinical characteristics were well balanced, with no significant differences in depressive symptom severity (PHQ-9: P=.56; QIDS-J: P=.74). No significant baseline differences were found between the AI-iCBT and iCBT groups, confirming the comparability of the 2 active interventions.

Figure 4 shows the flow of participants through the trial, including the numbers assessed for eligibility, randomized, allocated to each study arm (AI-iCBT, iCBT, control), completing follow-up assessments at week 7 and month 3, and included in the ITT analysis.

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) 2010 flow diagram of participant enrollment, allocation, follow-up, and analysis. AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy; iCBT: internet-based cognitive behavioral therapy; ITT: intention-to-treat.
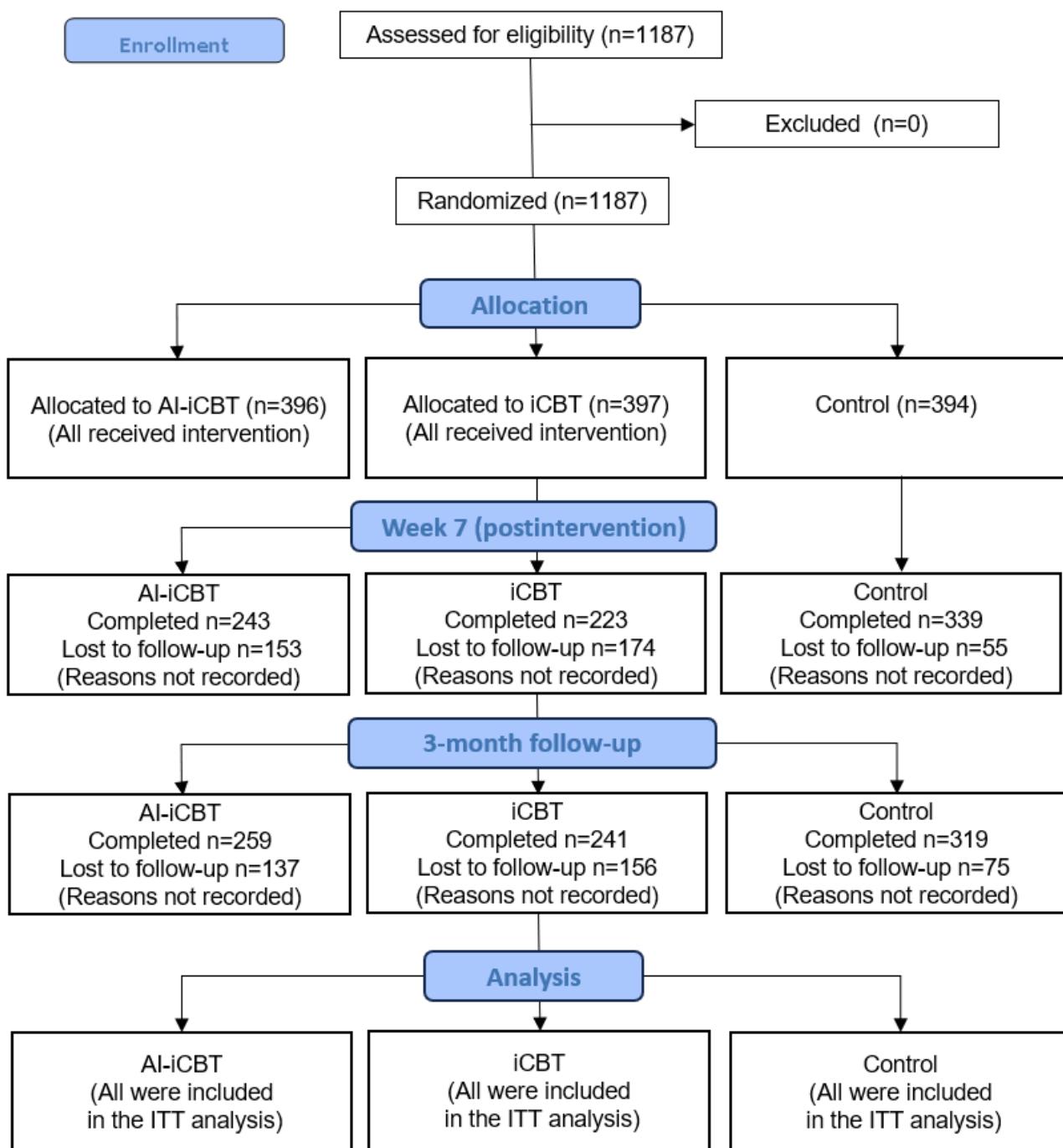
Table 1. Participants' characteristics.

Characteristic	Total (N=1187)	AI-iCBT[a] (n=396)	iCBT[b] (n=397)	Control (n=394)	P value (overall)	P value (AI-iCBT vs iCBT)[c]
Sex, n (%)					.96[d]	N/A[e]
Male	698 (58.8)	232 (58.6)	232 (58.4)	234 (59.4)		
Female	489 (41.2)	164 (41.4)	165 (41.6)	160 (40.6)		
Age (years)					.81[f]	N/A
Mean (SD)	43.5 (9.9)	43.6 (9.5)	43.2 (9.9)	43.6 (10.1)		
Median (IQR)	44 (36-52)	44 (36.8-51)	44 (36-52)	45 (36-52)		
Minimum-Maximum	20-60	20-60	20-60	20-60		
Marital status, n (%)					.28[d]	N/A
Married	665 (56)	226 (57.1)	219 (55.2)	220 (55.8)		
Divorced	76 (6.4)	29 (7.3)	17 (4.3)	30 (7.6)		
Bereaved	7 (0.6)	2 (0.5)	4 (1)	1 (0.3)		
Single	439 (37)	139 (35.1)	157 (39.5)	143 (36.3)		
Educational background, n (%)					.63[d]	N/A
Junior high school	7 (0.6)	3 (0.8)	1 (0.3)	3 (0.8)		
High school	239 (20.1)	87 (22)	84 (21.2)	68 (17.3)		
Junior college or technical	216 (18.2)	71 (17.9)	73 (18.4)	72 (18.3)		
University or postgraduate	725 (61.1)	235 (59.3)	239 (60.2)	251 (63.7)		
Employment status, n (%)					.78[d]	N/A
Working	972 (81.9)	320 (80.8)	324 (81.6)	328 (83.2)		
Unemployed (seeking)	79 (6.7)	27 (6.8)	30 (7.6)	22 (5.6)		
Unemployed (not seeking)	136 (11.5)	49 (12.4)	43 (10.8)	44 (11.2)		
Medical history, n (%)					.13[d]	N/A
No relevant history	918 (77.3)	316 (79.8)	296 (74.6)	306 (77.7)		
Ambulatory	258 (21.7)	74 (18.7)	97 (24.4)	87 (22.1)		
Hospitalized	11 (0.9)	6 (1.5)	4 (1)	1 (0.3)		
Mental history, n (%)					.93[d]	N/A
In treatment	129 (10.9)	46 (11.6)	44 (11.1)	39 (9.9)		
Treated	184 (15.5)	61 (15.4)	59 (14.9)	64 (16.2)		
No relevant history	874 (73.6)	289 (73)	294 (74.1)	291 (73.9)		
PHQ-9[g] score ≥10	428 (36.1)	142 (35.9)	143 (36)	143 (36.3)	.99[d]	N/A
Baseline scale score, mean (SD)						
PHQ-9	8.7 (5.2)	8.8 (5.2)	8.5 (5.2)	8.9 (5.1)	.56[f]	.41[f]
QIDS-J[h]	8.7 (4.9)	8.8 (4.9)	8.6 (4.9)	8.8 (4.7)	.74[f]	.49[f]
GAD-7[i]	6.0 (4.6)	6.0 (4.4)	5.9 (4.7)	6.2 (4.7)	.59[f]	.93[f]

[a]AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy.

[b]iCBT: internet-based cognitive behavioral therapy.

[c]P values represent pairwise comparisons between AI-iCBT and iCBT groups.

[d]P value is based on the chi-square test.

[e]N/A: not applicable.

[f]P value is based on ANOVA.

[g]PHQ-9: Patient Health Questionnaire-9.

[h]QIDS-J: Quick Inventory of Depressive Symptomatology-Japanese version.

[i]GAD-7: Generalized Anxiety Disorder-7.

Primary and Secondary Outcomes (ITT Population)

The primary outcome, the mean score on the PHQ-9, did not show statistically significant between-group differences compared with the control group at either week 7 or month 3 (AI-iCBT vs control: least squares mean difference –0.47, 95% CI –1.13 to 0.18; P=.16; Cohen d=–0.10; iCBT vs control: least squares mean difference –0.62, 95% CI –1.28 to 0.04; P=.07; Cohen d=–0.13; Table 2). No significant differences were observed between the AI-iCBT and iCBT groups. Nevertheless, both intervention groups showed significant within-group reductions from baseline at week 7 and month 3 (all P<.001), indicating that depressive symptoms improved over time in both groups.

Table 2. Primary outcome: mean Patient Health Questionnaire-9 scores at baseline, week 7, and month 3 (intention-to-treat population). Values are least squares (LS) means with 95% CIs estimated using a mixed model for repeated measures. Between-group comparisons are shown.

Time point	AI-iCBT[a] group, LS mean (95% CI)	iCBT[b] group, LS mean (95% CI)	Control group, LS mean (95% CI)	AI-iCBT vs control, Δ (95% CI)	iCBT vs control, Δ (95% CI)
Baseline	8.79 (8.47 to 9.11)	8.73 (8.41 to 9.05)	8.81 (8.48 to 9.13)	__[c]	—
Week 7	7.90 (7.49 to 8.31)	7.75 (7.33 to 8.18)	8.28 (7.94 to 8.63)	–0.37 (–1.02 to 0.29); P=.27	–0.46 (–1.13 to 0.21); P=.18
Month 3	7.37 (6.98 to 7.77)	7.17 (6.76 to 7.58)	7.86 (7.50 to 8.22)	–0.47 (–1.13 to 0.18), P=.16	–0.62 (–1.28– to 0.04); P=.07

[a]AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy.

[b]iCBT: internet-based cognitive behavioral therapy.

[c]Not available.

For the secondary binary outcome of PHQ-9 ≥10, the overall proportion decreased over time across all groups (Figure 5). At month 3, the proportion was numerically lower in the AI-iCBT group compared with the control group, but between-group differences were not statistically significant in the ITT analysis (Multimedia Appendix 2).

Figure 5. Secondary outcome: proportion of participants with Patient Health Questionnaire-9 (PHQ-9) scores ≥10 at baseline, week 7, and month 3 (intention-to-treat population). Estimated proportions and 95% CIs were derived from generalized linear mixed models with a logit link, including effects for group, time, and their interaction. AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy; iCBT: internet-based cognitive behavioral therapy.
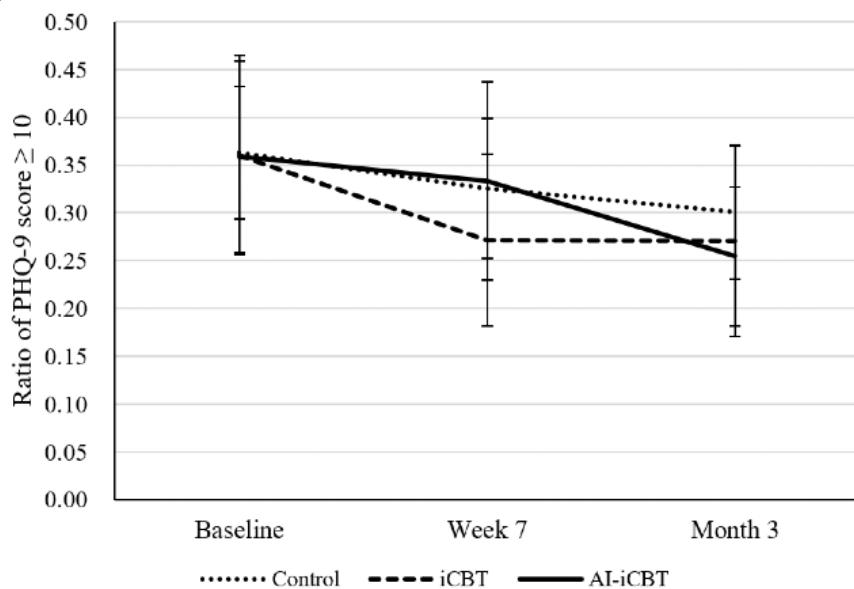
Similar patterns were observed for other secondary measures. QIDS-J and GAD-7 scores improved significantly within both intervention groups but without significant between-group differences. SDS scores showed modest reductions but did not significantly differ from control. Full secondary outcome results are provided in Multimedia Appendix 3.

Engagement and User Satisfaction

Overview

As illustrated in Figure 6, the CR exercise participation rate decreased significantly over time across both intervention groups (odds ratio [OR] 0.751, 95% CI 0.692-0.815; P<.001).

Participation began at only about half of participants in week 1 and declined further, dropping more steeply in the iCBT group, which fell to around 30% by week 6. In contrast, the AI-iCBT group retained somewhat higher engagement, remaining closer to the low 40% range by week 6, suggesting that AI support helped sustain participation over time. Between-group differences over time were examined using GLMM with a logit link, which showed a significant group × week interaction favoring AI-iCBT (OR 1.23, 95% CI 1.09-1.39; *P*<.001; Table 3). Analyses included all randomized participants in the intervention groups.

Figure 6. Engagement outcome: weekly participation rates in the artificial intelligence–augmented internet-based cognitive behavioral therapy (AI-iCBT) and internet-based cognitive behavioral therapy (iCBT) groups during weeks 1-6 (intention-to-treat population). Participation was defined as completion of at least 1 cognitive restructuring exercise per week. Error bars indicate 95% CIs.
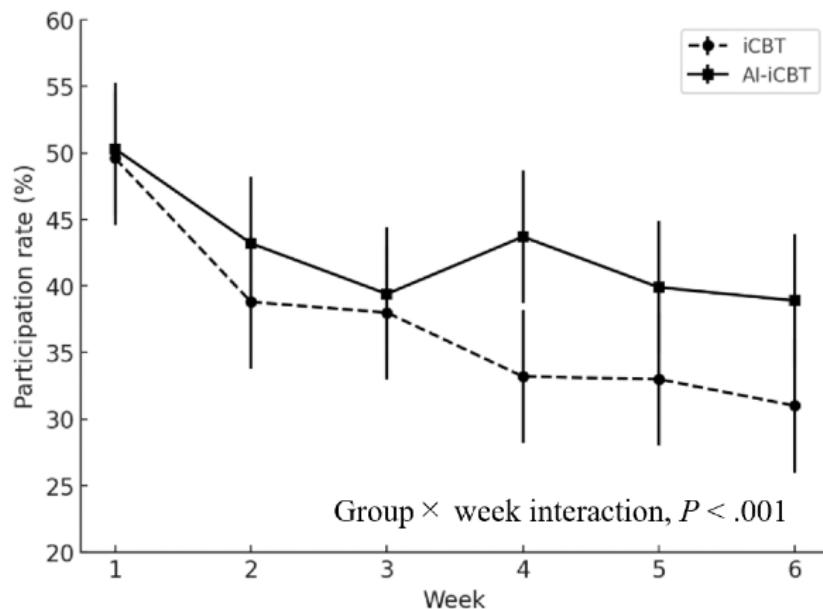
Table 3. Engagement outcome: mixed-effects logistic regression results for weekly cognitive restructuring exercise participation rates (intention-to-treat population). The generalized linear mixed model with a logit link included fixed effects for group, week (continuous and centered), and their interaction (group × week).

Effect	Reference	Odds ratio (95% CI)	*P* value
Group (AI-iCBT[a] vs iCBT[b])	iCBT	0.807 (0.370-1.758)	.59
Week (continuous, centered)	—[c]	0.751 (0.692-0.815)	<.001
Group × week	—	1.229 (1.090-1.386)	<.001

[a]AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy.

[b]iCBT: internet-based cognitive behavioral therapy.

[c]Not available.

User Satisfaction

Assessed at week 7 with the CSQ-8, averaged about 21 out of 32 points in both intervention groups, indicating a moderate to good level of satisfaction. No significant difference was observed between AI-iCBT and iCBT (Multimedia Appendix 4).

Exploratory Analysis of Engagement-Enhancing Factors

Activation of the empathy function was significantly associated with higher participation (OR 9.99, 95% CI 5.80-17.21; *P*<.001). In contrast, activation of the advisory function was not significantly associated with engagement (OR 2.37, 95% CI 0.96-5.83; *P*=.06). Detailed adjusted results are provided in Multimedia Appendix 5.

Exploratory EAS Analysis

In the exploratory EAS analysis, based on the ITT population, 312/396 (78.8%) in the AI-iCBT group, 312/397 (78.6%) in the iCBT group, and 317/394 (80.5%) in the control group had a baseline PHQ-9 score of ≥5. Among the ITT population, the proportion of participants who attended 3 or more sessions was 188/396 (47.5%) in the AI-iCBT group and 158/397 (39.8%) in the iCBT group. Furthermore, the proportion of participants with a baseline PHQ-9 score of ≥5 who attended 3 or more sessions (EAS) was 149/396 (37.6%) in the AI-iCBT group, 134/397 (33.8%) in the iCBT group, and 317/394 (80.5%) in the control group.

Mean PHQ-9 scores decreased significantly from baseline to week 7 in all 3 groups. At week 7, the iCBT group showed a statistically significant improvement compared with the control

group (Δ=–1.08, 95% CI –1.98 to –0.18; P=.02). However, this difference was not maintained at month 3. Full numerical results are provided in Multimedia Appendix 6. By contrast, the proportion of participants scoring ≥10 on the PHQ-9 decreased only in the iCBT group compared with the control at week 7 (Figure 7). At month 3, the AI-iCBT group showed a significantly lower proportion compared with control (Δ –0.15, 95% CI –0.30 to –0.01; P=.046), while the iCBT group did not differ significantly. The group × time interaction was significant (P=.008), indicating that the pattern of improvement differed between intervention groups (see Multimedia Appendix 7 for full numerical results).

Figure 7. Exploratory outcome: proportion of participants with Patient Health Questionnaire-9 (PHQ-9) scores of ≥10 in the efficacy analysis set population at baseline, week 7 (postintervention), and month 3 (follow-up). Error bars indicate 95% CIs. Asterisks represent P<.05 versus control. AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy; iCBT: internet-based cognitive behavioral therapy.
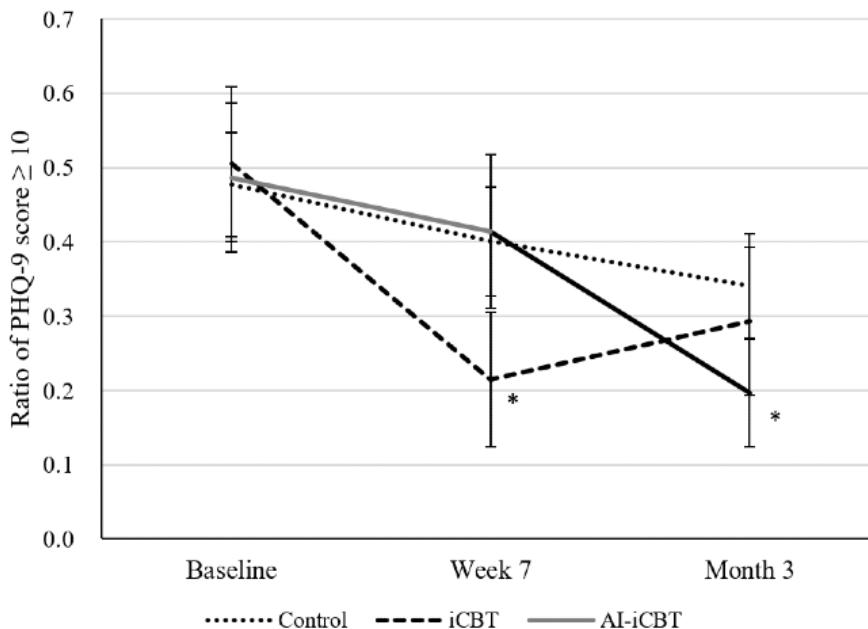
Sensitivity Analyses

In evaluating factors associated with low adherence, participants who attended <3 sessions were more likely to be male (65.1% vs 48.4%; P<.001), older (mean age 44.2, SD 9.44 vs mean age 41.7, SD 9.55 years; P=.001), and employed (287/341, 84.2% vs 212/283, 74.9%; P=.02), compared with those who attended ≥3 sessions. No significant differences were observed for marital status, educational background, medical history, or mental health history (Multimedia Appendix 8).

As a further sensitivity analysis, when applying a stricter definition of depression severity—PHQ-9 score of ≥10 plus at least 1 core symptom and SDS score of ≥10—62.7% (178/284) of participants with a PHQ-9 score of ≥10 at baseline met this definition (AI-iCBT: 44/70, 62.9%; iCBT: 41/71, 57.7%; control: 93/143, 65%), with no significant group differences (Multimedia Appendix 9).

Discussion

Overview

This study has several unique features. First, it directly compared a fully self-administered CR exercise delivered via iCBT, with and without AI-based NLP functionality, under a randomized (partially masked) design. Notably, the addition of AI led to a statistically significant improvement in engagement—an effect, to our knowledge, not previously documented. As AI-based interventions have become increasingly sophisticated and deeply integrated into intervention programs, disentangling the specific contribution of AI has become difficult. In particular, establishing control conditions that differ only in the presence or absence of AI functionality requires substantial resources, and prior studies have therefore often relied on heterogeneous control conditions. By applying a more robust design—feasible in part because the technology was still in a transitional phase—this trial provides new insights into how AI may enhance fully self-administered iCBT and offers a timely perspective for advancing scalable mental health care.

Self-administered interventions are known to be modestly to moderately effective for depression [15,23], but adherence remains a major limitation [15,18,19,21,55,63]. Systematic reviews indicate that approximately one-third to one-half of participants drop out before completing the program [19,64]. In this context, the engagement improvement observed in this study represents a potential step toward overcoming this barrier.

With regard to clinical effectiveness, no additional benefits of NLP feedback were observed in the ITT population. Recent evidence has reported that greater antidepressant effects are associated with lower dropout rates [15,20,65]. In contrast, although no between-group differences in antidepressant effects were found here, the addition of AI feedback was associated with a statistically significant increase in adherence. This suggests a novel engagement-enhancing mechanism, distinct from the traditionally assumed link between larger clinical effects and lower dropout rates.

Exploratory analyses further indicated that the "empathic function" of AI feedback was significantly associated with

improved adherence, whereas the advisory function showed no significant effect. Participants who received empathic responses during their first exercise subsequently demonstrated higher adherence. While self-disclosure was not directly measured, the sense of being supported may have facilitated persistence. These findings align with prior evidence that empathic conversational agents and chatbots support therapeutic alliance and sustained engagement [66-69]. Research in behavioral change has likewise shown that AI-mediated feedback can promote sustained self-management [70], supporting the plausibility of these findings. Such effects of human-AI interaction may have been less visible in prior studies using heterogeneous control conditions but became evident here through the structured randomized design.

Regarding antidepressant effects, no significant between-group differences were observed in the ITT population. In the EAS, results—while requiring cautious interpretation—indicated that at week 7 only the iCBT group showed significant improvement in both the mean PHQ-9 score and the proportion of participants with a PHQ-9 score of ≥10 compared with the control group, whereas the AI-iCBT group did not.

This suggests that the AI function may have attenuated or failed to enhance short-term antidepressant effects. However, this short-term benefit in the iCBT group disappeared at long-term follow-up. For the continuous outcome in particular, the short-term difference was –1.1 points on the PHQ-9, below the minimal clinically important difference (approximately 3 points) [35,71], suggesting limited clinical significance.

By contrast, the short-term dichotomous outcome in the iCBT group represented about a 29% reduction in the proportion of PHQ-9 scores of ≥10 cases relative to the control group. This implies that part of the potential benefit may not have been realized when AI was introduced. At long-term follow-up (month 3), however, only the AI-iCBT group showed a significant reduction of about 15% compared with the control group. These findings highlight the absence of the expected short-term effect in AI-iCBT and the unique long-term effect observed only in AI-iCBT.

The fact that AI-iCBT ultimately demonstrated an effect at long-term follow-up is noteworthy. Although exploratory, this suggests a potential contribution of AI-iCBT in reducing the proportion of participants exceeding a clinically significant severity threshold. One possible explanation is that approaches emphasizing empathy as a core therapeutic skill—such as interpersonal psychotherapy or family therapies—often show slower onset but more enduring gains compared with CBT, lasting well beyond the end of treatment [72-77]. It is possible that the empathy-related function of AI contributed in a similar way, although the underlying mechanisms remain unclear.

The EAS, however, was more restrictive than the ITT population. Among ITT participants with a baseline PHQ-9 score of ≥5, only 149/312 (47.8%) in the AI-iCBT group and 134/312 (42.9%) in the iCBT group attended at least 3 sessions. Furthermore, although a PHQ-9 score of ≥10 is widely recognized as a proxy for "major depression equivalent" in research [33,39], concerns have been raised that it may not be sufficient for diagnostic purposes and may contribute to

overdiagnosis [56,78,79]. As a sensitivity analysis, therefore, we used a stricter definition requiring a PHQ-9 score of ≥10 plus at least 1 core symptom (depressed mood or anhedonia) [56,57], together with an SDS score of ≥10 as an indicator of functional impairment [49,58-60]. Results confirmed that only about 62.7% (178/284) of participants who met the PHQ-9 score of ≥10 at baseline also met this stricter definition, highlighting the importance of cautious interpretation (Multimedia Appendix 9).

Taken together, regardless of whether it corresponds to major depression, the finding of a significant reduction in the proportion of participants with clinically meaningful depressive states (PHQ-9 ≥10) compared with the control group may have clinical significance, particularly given the fully self-help nature of the program. From a public health perspective, such a difference could also carry implications for the scalability of self-help programs that do not require therapist involvement.

AI communication, including generative AI, continues to advance rapidly. However, the development of appropriate control programs has often been constrained by logistical and cost-related factors, limiting opportunities to rigorously investigate the antidepressant and anxiolytic effects of AI. Beyond psychiatry, maximizing the effectiveness of self-administered interventions while enhancing engagement remains a critical challenge across health care and welfare domains. This study represents a step toward addressing this challenge.

Limitations

First, most participants were recruited from a research panel with high affinity for digital interventions. Only 10.9% (129/1187) were actual users of mental health services, which is consistent with the national average in Japan, but caution is required in generalizing these findings to broader populations. Second, high dropout rates were observed, with only 37.6% (149/396) of the AI-iCBT group and 33.8% (134/397) of the iCBT group in the ITT population meeting EAS criteria. Additional analyses indicated that low adherence was associated with being male, older, and employed (Multimedia Appendix 8). Consistent with recent studies, time constraints [22], particularly among employed men [21] and older adults [20,21], were confirmed as key barriers to engagement. Third, although the AI-iCBT group consistently showed 5%-10% higher adherence rates than the iCBT group throughout the trial, both groups had already dropped by half to 50% (396/793) at the very first session and continued to decline over time, remaining at only 30%-40%.

Fourth, missing data were substantial, and MMRM and generalized estimating equations were applied to minimize bias. However, these approaches assume data are missing at random. In this study, since attrition occurred according to participant attributes, the possibility of missing not at random cannot be ruled out, and estimates may remain biased. Fifth, a significant group × time interaction was observed in adherence in the AI feedback group, suggesting a potential role of AI in improving engagement. However, this conclusion is based on a single trial and requires replication in different populations and designs, as well as further elucidation of the underlying mechanisms.

Sixth, due to technical issues, session-by-session data on depressive symptoms (Overall Anxiety Severity and Impairment Scale) [80] and anxiety (Overall Depression Severity and Impairment Scale) [81] were lost, precluding more detailed analyses. Future studies should implement automatic data saving and backup systems to prevent such loss. Overall antidepressant effects were small. In addition to limitations of the program itself, this may reflect a ceiling effect due to the predominance of participants with mild depression. Several meta-analyses [23,82,83] report that treatment effects may be less pronounced in cases with mild baseline depression compared to moderate or severe cases. Another limitation is related to our stratified randomization. While stratification by age, sex, and baseline PHQ-9 severity increased internal validity by balancing key prognostic factors, it may also restrict the generalizability of our findings to populations with different distributions of these characteristics.

Despite these challenges, this study provides valuable insights into the potential of AI-enhanced self-help digital interventions, particularly in relation to participant behavior dynamics. Importantly, no major adverse events were reported, underscoring the safety of this innovative approach and its potential to significantly advance mental health care practices.

Data Availability

The data used in this study belong to the Research Institute of Economy, Trade and Industry and can be obtained from the institute upon reasonable request.

Authors' Contributions

MS contributed to conceptualization, investigation, methodology, project administration, resources, software, validation, visualization, and writing—including original draft preparation, review, and editing. YS contributed to conceptualization, funding acquisition, investigation, methodology, project administration, resources, and writing—review and editing. SH contributed to data curation, formal analysis, validation, visualization, and writing—review and editing. MK and HY contributed to secondary analysis and writing—review and editing. NW contributed to conceptualization, investigation, methodology, supervision, validation, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist.
[PDF File (Adobe PDF File), 340 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Secondary outcome: ratio of participants with Patient Health Questionnaire-9 scores ≥10 (intention-to-treat population).
[DOCX File , 17 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Secondary outcomes: mean score of Quick Inventory of Depressive Symptomatology-Japanese version, Generalized Anxiety Disorder-7, and Sheehan Disability Scale (intention-to-treat population).
[DOCX File , 19 KB-Multimedia Appendix 3]

Multimedia Appendix 4

Satisfaction outcome: Client Satisfaction Questionnaire-8 total scores at week 7 (intention-to-treat population, intervention groups only).
[DOCX File , 17 KB-Multimedia Appendix 4]

Multimedia Appendix 5

Associations between artificial intelligence feedback functions and exercise attendance (artificial intelligence–augmented internet-based cognitive behavioral therapy and internet-based cognitive behavioral therapy participants, weeks 2–6).
[DOCX File , 20 KB-Multimedia Appendix 5]

Multimedia Appendix 6

Exploratory outcome: mean Patient Health Questionnaire-9 scores (efficacy analysis set population).
[DOCX File , 18 KB-Multimedia Appendix 6]

Multimedia Appendix 7

Exploratory outcome: ratio of participants with Patient Health Questionnaire-9 scores ≥10 (efficacy analysis set population).
[DOCX File , 18 KB-Multimedia Appendix 7]

Multimedia Appendix 8

Associated factors of low adherence in the efficacy analysis set.
[DOCX File , 20 KB-Multimedia Appendix 8]

Multimedia Appendix 9

Baseline prevalence of major depression proxies in the efficacy analysis set population.
[DOCX File , 18 KB-Multimedia Appendix 9]

References

1. GBD 2017 DiseaseInjury IncidencePrevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. Lancet. 2018;392(10159):1789-1858. [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Arias D, Saxena S, Verguet S. Quantifying the global burden of mental disorders and their economic value. EClinicalMedicine. 2022;54:101675. [FREE Full text] [doi: 10.1016/j.eclinm.2022.101675] [Medline: 36193171]
3. Hohls JK, König H-H, Quirke E, Hajek A. Anxiety, depression and quality of life-a systematic review of evidence from longitudinal observational studies. Int J Environ Res Public Health. 2021;18(22):12022. [FREE Full text] [doi: 10.3390/ijerph182212022] [Medline: 34831779]
4. König H, König H-H, Konnopka A. The excess costs of depression: a systematic review and meta-analysis. Epidemiol Psychiatr Sci. 2019;29:e30. [FREE Full text] [doi: 10.1017/S2045796019000180] [Medline: 30947759]
5. WHO. Mental health at work. World Health Organization. Geneva.; 2022. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work [accessed 2025-12-26]
6. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. Health Aff (Millwood). 2009;28(3):w490-w501. [doi: 10.1377/hlthaff.28.3.w490] [Medline: 19366722]
7. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet. 2007;370(9590):841-850. [FREE Full text] [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
8. Wainberg ML, Lu FG, Riba MB. Global mental health. Acad Psychiatry. 2016;40(4):647-649. [doi: 10.1007/s40596-016-0577-0] [Medline: 27259490]
9. Edge D, Watkins ER, Limond J, Mugadza J. The efficacy of self-guided internet and mobile-based interventions for preventing anxiety and depression: a systematic review and meta-analysis. Behav Res Ther. 2023;164:104292. [FREE Full text] [doi: 10.1016/j.brat.2023.104292] [Medline: 37003138]
10. Psychological Interventions Implementation Manual: Integrating Evidence-Based Psychological Interventions into Existing Services. Geneva. World Health Organization; 2024.
11. WHO. Consolidated telemedicine implementation guide. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/9789240059184 [accessed 2025-12-26]

12. Binder P, Hjeltnes A. Mindfulness in psychotherapy and society—the need for combining enthusiasm and critical inquiry. Couns and Psychother Res. 2021;21(2):247-250. [FREE Full text] [doi: 10.1002/capr.12384]

13. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. BMJ Open. 2015;5(11):e008581. [FREE Full text] [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]

14. Andersson G, Carlbring P, Rozental A. Response and remission rates in internet-based cognitive behavior therapy: an individual patient data meta-analysis. Front Psychiatry. 2019;10:749. [doi: 10.3389/fpsyt.2019.00749] [Medline: 31708813]

15. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. JAMA Psychiatry. 2021;78(4):361-371. [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.4364] [Medline: 33471111]

16. Saad A, Bruno D, Camara B, D'Agostino J, Bolea-Alamanac B. Self-directed technology-based therapeutic methods for adult patients receiving mental health services: systematic review. JMIR Ment Health. 2021;8(11):e27404. [FREE Full text] [doi: 10.2196/27404] [Medline: 34842556]

17. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry. 2019;76(7):700-707. [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]

18. Seittu HA, Falk T, Bhatnagar K, Saarni SE. Therapists' role in patient adherence to internet-based cognitive behavioral therapy: qualitative study. J Med Internet Res. 2025;27:e71852. [FREE Full text] [doi: 10.2196/71852] [Medline: 40929730]

19. Treanor CJ, Kouvonen A, Lallukka T, Donnelly M. Acceptability of computerized cognitive behavioral therapy for adults: umbrella review. JMIR Ment Health. 2021;8(7):e23091. [FREE Full text] [doi: 10.2196/23091] [Medline: 34255714]

20. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an internet intervention for depression. J Affect Disord. 2018;229:443-449. [doi: 10.1016/j.jad.2017.12.028] [Medline: 29331706]

21. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med. 2015;45(13):2717-2726. [FREE Full text] [doi: 10.1017/S0033291715000665] [Medline: 25881626]

22. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med. 2016;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]

23. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry. 2017;74(4):351-359. [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]

24. Kemmeren L, van Schaik A, Draisma S, Kleiboer A, Riper H, Smit J. Effectiveness of blended cognitive behavioral therapy versus treatment as usual for depression in routine specialized mental healthcare: E-COMPARED trial in the Netherlands. Cogn Ther Res. 2023;47(3):386-398. [doi: 10.1007/s10608-023-10363-y]

25. Mathiasen K, Andersen TE, Lichtenstein MB, Ehlers LH, Riper H, Kleiboer A, et al. The clinical effectiveness of blended cognitive behavioral therapy compared with face-to-face cognitive behavioral therapy for adult depression: randomized controlled noninferiority trial. J Med Internet Res. 2022;24(9):e36577. [FREE Full text] [doi: 10.2196/36577] [Medline: 36069798]

26. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol. 2018;14:91-118. [doi: 10.1146/annurev-clinpsy-032816-045037] [Medline: 29401044]

27. Nie J, Shao H, Fan Y, Shao Q, You H, Preindl M, et al. LLM-based conversational AI therapist for daily functioning screening and psychotherapeutic intervention via everyday smart devices. arXiv. Preprint posted online on March 16, 2024. 2025. [FREE Full text] [doi: 10.1145/3712299]

28. Malgaroli M, Hull TD, Zech JM, Althoff T. Natural language processing for mental health interventions: a systematic review and research framework. Transl Psychiatry. 2023;13(1):309. [FREE Full text] [doi: 10.1038/s41398-023-02592-2] [Medline: 37798296]

29. Le Glaz A, Haralambous Y, Kim-Dufor D-H, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res. 2021;23(5):e15708. [doi: 10.2196/15708] [Medline: 33944788]

30. Sheehan K, Bhatti PK, Yousuf S, Rosenow W, Roehler DR, Hazekamp C, et al. Natural language processing applied to mental illness detection: a narrative review. NPJ Digit Med. 2022;5(1):46. [FREE Full text] [doi: 10.1186/s12889-022-13177-x] [Medline: 35449046]

31. Choudhury MD, Pendse SR, Kumar N. Benefits and harms of large language models in digital mental health. arXiv. Preprint posted online on November 7, 2021. 2021. [FREE Full text]

32. Villarreal-Zegarra D, Reategui-Rivera CM, García-Serna J, Quispe-Callo G, Lázaro-Cruz G, Centeno-Terrazas G, et al. Self-administered interventions based on natural language processing models for reducing depressive and anxiety symptoms: systematic review and meta-analysis. JMIR Ment Health. 2024;11:e59560. [FREE Full text] [doi: 10.2196/59560] [Medline: 39167795]

33. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-613. [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

34. Muramatsu K, Miyaoka H, Kamijima K, Muramatsu Y, Tanaka Y, Hosaka M, et al. Performance of the Japanese version of the Patient Health Questionnaire-9 (J-PHQ-9) for depression in primary care. Gen Hosp Psychiatry. 2018;52:64-69. [doi: 10.1016/j.genhosppsych.2018.03.007] [Medline: 29698880]

35. Bauer-Staeb C, Kounali D-Z, Welton NJ, Griffith E, Wiles NJ, Lewis G, et al. Effective dose 50 method as the minimal clinically important difference: evidence from depression trials. J Clin Epidemiol. 2021;137:200-208. [FREE Full text] [doi: 10.1016/j.jclinepi.2021.04.002] [Medline: 33892086]

36. Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, et al. How much change is enough? Evidence from a longitudinal study on depression in UK primary care. Psychol Med. 2022;52(10):1875-1882. [FREE Full text] [doi: 10.1017/S0033291720003700] [Medline: 33138872]

37. So M, Sekizawa Y, Yamaguchi Y. A randomised controlled trial investigating the clinical and cost-effectiveness of peer enhanced-computerised cognitive depression. KAKEN. 2015. URL: https://kaken.nii.ac.jp/en/grant/KAKENHI-PROJECT-26350862/ [accessed 2025-12-19]

38. Inagaki M, Ohtsuki T, Yonemoto N, Kawashima Y, Saitoh A, Oikawa Y, et al. Validity of the Patient Health Questionnaire (PHQ)-9 and PHQ-2 in general internal medicine primary care at a Japanese rural hospital: a cross-sectional study. Gen Hosp Psychiatry. 2013;35(6):592-597. [doi: 10.1016/j.genhosppsych.2013.08.001] [Medline: 24029431]

39. Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. CMAJ. 2012;184(3):E191-E196. [FREE Full text] [doi: 10.1503/cmaj.110829] [Medline: 22184363]

40. Cuijpers P, Karyotaki E, Ciharova M, Miguel C, Noma H, Furukawa TA. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. Acta Psychiatr Scand. 2021;144(3):288-299. [doi: 10.1111/acps.13335] [Medline: 34107050]

41. Fava M. Depression with physical symptoms: treating to remission. J Clin Psychiatry. 2003;64 Suppl 7:24-28. [Medline: 12755649]

42. Keller MB, Lavori PW, Mueller TI, Endicott J, Coryell W, Hirschfeld RM, et al. Time to recovery, chronicity, and levels of psychopathology in major depression. A 5-year prospective follow-up of 431 subjects. Arch Gen Psychiatry. 1992;49(10):809-816. [doi: 10.1001/archpsyc.1992.01820100053010] [Medline: 1417434]

43. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol Psychiatry. 2003;54(5):573-583. [doi: 10.1016/s0006-3223(02)01866-8] [Medline: 12946886]

44. Fujisawa D, Nakagawa A, Tajima M, Ono Y. Development of Japanese version of QIDS-SR (self-report). Jpn J Stress Sci. 2010;25(1):43-52. [FREE Full text]

45. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med. 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

46. Muramatsu K, Muramatsu Y, Miyaoka H, Fuse K, Yoshimine F, Hosaka M. Validation and utility of a Japanese version of the GAD-7. In: Panminerva Medica. 2009. Presented at: 20th World Congress on Psychosomatic Medicine; August 23-28, 2009:79; Torino, Italy.

47. Sheehan KH, Sheehan DV. Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale. Int Clin Psychopharmacol. 2008;23(2):70-83. [doi: 10.1097/YIC.0b013e3282f2b4d6] [Medline: 18301121]

48. Yoshida T, Otsubo T, Tsuchida H, Wada R, Ueshima K, Fukui A. Reliability and validity of the Japanese version of the Sheehan Disability Scale. Jpn J Clin Psychopharmacol. 2004;7(10):1645-1653.

49. Soares CN, Zhang M, Boucher M. Categorical improvement in functional impairment in depressed patients treated with desvenlafaxine. CNS Spectr. 2019;24(3):322-332. [FREE Full text] [doi: 10.1017/S1092852917000633] [Medline: 29140227]

50. Attkisson CC, Zwick R. The Client Satisfaction Questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann. 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]

51. Tachimori H, Ito H. Reliability and validity of the Japanese version of the Client Satisfaction Questionnaire. Seishin Igaku (Clin Psychiatry). 1999;41(7):711-717. [doi: 10.11477/mf.1405905056]

52. Wong SYS, Sun YY, Chan ATY, Leung MKW, Chao DVK, Li CCK, et al. Treating subthreshold depression in primary care: a randomized controlled trial of behavioral activation with mindfulness. Ann Fam Med. 2018;16(2):111-119. [FREE Full text] [doi: 10.1370/afm.2206] [Medline: 29531101]

53. Harrer M, Sprenger AA, Illing S, Adriaanse MC, Albert SM, Allart E, et al. Psychological intervention in individuals with subthreshold depression: individual participant data meta-analysis of treatment effects and moderators. Br J Psychiatry. 2025:1-14. [FREE Full text] [doi: 10.1192/bjp.2025.56] [Medline: 40365980]

54. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res. 2013;15(10):e231. [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]

55. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res. 2011;13(3):e52. [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]

56. Levis B, Benedetti A, Thombs BD, DEPRESsion Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. BMJ. 2019;365:l1476. [FREE Full text] [doi: 10.1136/bmj.l1476] [Medline: 30967483]

57. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals. 2002;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]

58. Lam RW, Michalak EE, Yatham LN. A new Clinical Rating Scale for work absence and productivity: validation in patients with major depressive disorder. BMC Psychiatry. 2009;9:78. [FREE Full text] [doi: 10.1186/1471-244X-9-78] [Medline: 19958540]

59. Luciano JV, Bertsch J, Salvador-Carulla L, Tomás JM, Fernández A, Pinto-Meza A, et al. Factor structure, internal consistency and construct validity of the Sheehan Disability Scale in a Spanish primary care sample. J Eval Clin Pract. 2010;16(5):895-901. [doi: 10.1111/j.1365-2753.2009.01211.x] [Medline: 20626541]

60. Sheehan DV, Harnett-Sheehan K, Raj BA. The measurement of disability. Int Clin Psychopharmacol. 1996;11 Suppl 3:89-95. [doi: 10.1097/00004850-199606003-00015] [Medline: 8923116]

61. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332. [FREE Full text] [doi: 10.1136/bmj.c332] [Medline: 20332509]

62. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res. 2011;13(4):e126. [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]

63. Eysenbach G. The law of attrition. J Med Internet Res. 2005;7(1):e11. [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

64. Koelen JA, Vonk A, Klein A, de Koning L, Vonk P, de Vet S, 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. [FREE Full text] [doi: 10.1016/j.cpr.2022.102179] [Medline: 35763975]

65. Kambeitz-Ilankovic L, Rzayeva U, Völkel L, Wenzel J, Weiske J, Jessen F, et al. A systematic review of digital and face-to-face cognitive behavioral therapy for depression. NPJ Digit Med. 2022;5(1):144. [FREE Full text] [doi: 10.1038/s41746-022-00677-8] [Medline: 36109583]

66. 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;4:847991. [FREE Full text] [doi: 10.3389/fdgth.2022.847991] [Medline: 35480848]

67. 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;4(2):e19. [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]

68. 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. [FREE Full text] [doi: 10.2196/12106] [Medline: 30470676]

69. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns. 2005;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]

70. Prochaska JJ, Vogel EA, 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;23(3):e24850. [FREE Full text] [doi: 10.2196/24850] [Medline: 33755028]

71. Lynch CP, Cha EDK, Jenkins NW, Parrish JM, Mohan S, Jadczak CN, et al. The minimum clinically important difference for Patient Health Questionnaire-9 in minimally invasive transforaminal interbody fusion. Spine (Phila Pa 1976). 2021;46(9):603-609. [doi: 10.1097/BRS.0000000000003853] [Medline: 33290370]

72. Fairburn CG, Norman PA, Welch SL, O'Connor ME, Doll HA, Peveler RC. A prospective study of outcome in bulimia nervosa and the long-term effects of three psychological treatments. Arch Gen Psychiatry. 1995;52(4):304-312. [doi: 10.1001/archpsyc.1995.03950160054010] [Medline: 7702447]

73. Agras WS, Walsh T, Fairburn CG, Wilson GT, Kraemer HC. A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa. Arch Gen Psychiatry. 2000;57(5):459-466. [doi: 10.1001/archpsyc.57.5.459] [Medline: 10807486]

74. Carter FA, Jordan J, McIntosh VVW, Luty SE, McKenzie JM, Frampton CMA, et al. The long-term efficacy of three psychotherapies for anorexia nervosa: a randomized, controlled trial. Int J Eat Disord. 2011;44(7):647-654. [doi: 10.1002/eat.20879] [Medline: 21997429]

75. Markowitz JC, Petkova E, Neria Y, Van Meter PE, Zhao Y, Hembree E, et al. Is exposure necessary? A randomized clinical trial of interpersonal psychotherapy for PTSD. Am J Psychiatry. 2015;172(5):430-440. [FREE Full text] [doi: 10.1176/appi.ajp.2014.14070908] [Medline: 25677355]

76. Bighelli I, Rodolico A, García-Mieres H, Pitschel-Walz G, Hansen WP, Schneider-Thoma J, et al. Psychosocial and psychological interventions for relapse prevention in schizophrenia: a systematic review and network meta-analysis. Lancet Psychiatry. 2021;8(11):969-980. [doi: 10.1016/j.envres.2021.112166] [Medline: 34619129]

77. Lemmens LHJM, van Bronswijk SC, Peeters FPML, Arntz A, Roefs A, Hollon SD, et al. Interpersonal psychotherapy versus cognitive therapy for depression: how they work, how long, and for whom-key findings from an RCT. Am J Psychother. 2020;73(1):8-14. [doi: 10.1176/appi.psychotherapy.20190030] [Medline: 32122161]

78. Mitchell AJ, Yadegarfar M, Gill J, Stubbs B. Case finding and screening clinical utility of the Patient Health Questionnaire (PHQ-9 and PHQ-2) for depression in primary care: a diagnostic meta-analysis of 40 studies. BJPsych Open. 2016;2(2):127-138. [FREE Full text] [doi: 10.1192/bjpo.bp.115.001685] [Medline: 27703765]

79. Levis B, Bhandari PM, Neupane D, Fan S, Sun Y, He C, et al. Depression Screening Data (DEPRESSD) PHQ Group. Data-driven cutoff selection for the Patient Health Questionnaire-9 depression screening tool. JAMA Netw Open. 2024;7(11):e2429630. [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.29630] [Medline: 39576645]

80. Bentley KH, Gallagher MW, Carl JR, Barlow DH. Development and validation of the Overall Depression Severity and Impairment Scale. Psychol Assess. 2014;26(3):815-830. [doi: 10.1037/a0036216] [Medline: 24708078]

81. Norman SB, Cissell SH, Means-Christensen AJ, Stein MB. Development and validation of an Overall Anxiety Severity and Impairment Scale (OASIS). Depress Anxiety. 2006;23(4):245-249. [doi: 10.1002/da.20182] [Medline: 16688739]

82. Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Huibers MJH. How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. World Psychiatry. 2016;15(3):245-258. [FREE Full text] [doi: 10.1002/wps.20346] [Medline: 27717254]

83. Mercorio A, Zizolfi B, Barbuto S, Danzi R, Di Spiezio Sardo A, Moawad G, et al. Three-dimensional imaging reconstruction and laparoscopic robotic surgery: a winning combination for a complex case of multiple myomectomy. Fertil Steril. 2023;120(1):202-204. [FREE Full text] [doi: 10.1016/j.fertnstert.2023.04.015] [Medline: 37085096]

Abbreviations

AI: artificial intelligence
AI-iCBT: artificial intelligence–augmented internet-based cognitive behavioral therapy
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
CR: cognitive restructuring
CSQ-8: Client Satisfaction Questionnaire-8
EAS: efficacy analysis set
GAD-7: Generalized Anxiety Disorder-7
GLMM: generalized linear mixed models
iCBT: internet-based cognitive behavioral therapy
ITT: intention-to-treat
MMRM: mixed-effects model for repeated measures
NLP: natural language processing
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
QIDS-J: Quick Inventory of Depressive Symptomatology-Japanese version
SDS: Sheehan Disability Scale

XSL•FO
RenderX

XSL•FO
RenderX