

Does AI Communication Truly Bring Benefits to CBT for Depression: A randomized, partial-double blind, parallel-group, comparative, exploratory clinical trial

Mirai So, Yoichi Sekizawa, Sora Hashimoto, Masami Kashimura, Hajime Yamakage, Norio Watanabe

Submitted to: Journal of Medical Internet Research
on: May 03, 2025

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

Table of Contents

Original Manuscript.....	5
---------------------------------	----------

Preprint
JMIR Publications

Does AI Communication Truly Bring Benefits to CBT for Depression: A randomized, partial-double blind, parallel-group, comparative, exploratory clinical trial

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

¹Department of Psychiatry Tokyo Dental College Tokyo JP

²Research Institute of Economy, Trade and Industry Tokyo JP

³United Health Communication Co., Ltd. Tokyo JP

⁴Department of Psychology Faculty of Human Sciences Tokai University Ibaraki JP

⁵Department of Medical Statistics Satista Co., Ltd. Kyoto JP

⁶Department of Psychiatry Soseikai General Hospital Kyoto JP

*these authors contributed equally

Corresponding Author:

Mirai So MD, PhD, MBA

Department of Psychiatry

Tokyo Dental College

2-9-18 Misakicho, Chiyoda-ku

Tokyo

Tokyo

JP

Abstract

Background: Depression is a leading cause of disability worldwide, significantly impacting quality of life and imposing substantial financial burdens. Technology-delivered self-help interventions, including internet-based cognitive behavioral therapy (iCBT), have emerged as scalable and cost-effective solutions, particularly for underserved populations. However, fully self-administered interventions often struggle with limited effectiveness and low engagement. Recently, Natural Language Processing (NLP) has been applied to enhance treatment effectiveness and adherence by improving user experience through personalized interactions. Despite promising applications, previous studies have primarily compared NLP-based interventions to passive controls or heterogeneous alternatives, making it difficult to isolate NLP's specific impact.

Objective: This study aimed to evaluate the effectiveness of NLP-supported iCBT in a fully self-administered setting by directly comparing it with a standard iCBT program (without NLP) and a waitlist control, within a structured exploratory randomized controlled trial design.

Methods: We implemented a randomized, double-blind, parallel-group exploratory trial to compare fully self-administered AI-supported iCBT (iCBT with NLP advisory or empathetic feedback as therapeutic contribution) and standard iCBT (without NLP feedback), both using an otherwise identical iCBT program, along with a waitlist control group. Participants (N = 1,187) were recruited online, and those meeting inclusion criteria (PHQ-9 ≥ 5) were randomly assigned to AI-iCBT, standard iCBT, or waitlist groups. The primary outcome was the proportion of participants with PHQ-9 scores ≥ 10 at Week 7 and Month 3. Secondary outcomes included changes in depressive and anxiety symptoms, functional impairment, adherence, and user satisfaction. The analysis followed the modified intention-to-treat (mITT) principle. Mixed-effects models for repeated measures (MMRM) were used for statistical analysis, adjusting for baseline imbalances and stratification factors (age, gender, and baseline PHQ-9 score).

Results: At Week 7, the iCBT group showed the most significant reduction in the proportion of participants with PHQ-9 scores ≥ 10 (21.4%), followed by AI-iCBT (41.4%) and waitlist (40.0%). However, by Month 3, AI-iCBT achieved a significant reduction to 19.7% ($p = 0.046$ vs. control), while the iCBT group rebounded to 29.3% ($p = 0.413$ vs. control). This suggests that AI-iCBT may provide more sustained antidepressant effects. Adherence trends showed AI-iCBT participants had lower early participation rates but better retention after Week 4 ($p = 0.026$ for time-by-group interaction).

Conclusions: To our knowledge, this is the first randomized trial directly isolating NLP's impact on depression in fully self-

administered iCBT. Results suggest that while standard iCBT may provide faster short-term symptom relief, AI-enhanced iCBT led to more sustained improvements and a greater reduction in the number of participants with moderate to severe depression over time. These findings highlight the potential of NLP to optimize digital mental health interventions, although further research is needed to refine its implementation and maximize its long-term benefits. Clinical Trial: UMIN00001922

(JMIR Preprints 03/05/2025:76902)

DOI: <https://doi.org/10.2196/preprints.76902>

Preprint Settings

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

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

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

Only make the preprint title and abstract visible.

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

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

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

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

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/76902>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://preprints.jmir.org/preprint/76902>

Original Manuscript

Does AI communication truly bring benefits to CBT for depression?**□A randomized, partial double-blind, parallel-group exploratory trial***A brief title:*

Does AI communication truly brings benefits to CBT for depression–fact or fiction?

Author name(s):

Mirai So

Yoichi Sekizawa

Sora Hashimoto

Masami Kashimura

Hajime Yamakage

Norio Watanabe

Affiliation(s):

Department of Psychiatry, Tokyo Dental College, Tokyo, Japan (M So, MD); Research Institute of Economy, Trade and Industry, Tokyo, Japan (Y Sekizawa, MA); United Health Communication Co., Ltd., Tokyo, Japan (S Hashimoto, MA); Department of Psychology, Tokiwa University, Ibaraki, Japan (M Kashimura, PhD); Department of Medical Statistics, Satista Co., Ltd., Kyoto, Japan (H Yamakage, M.Eng.); and Department of Psychiatry, Soseikai General Hospital, Kyoto, Japan (N Watanabe, MD)

Correspondence to:

Mirai So, MD

Department of Psychiatry, Tokyo Dental College, 2-9-18, Misakicho, Chiyoda-ku, Tokyo, 101-0061, Japan

Phone: 81-47-322-0151

Email: somirai@tdc.ac.jp

Abstract 450 words**Background**

Depression is a leading cause of disability worldwide, significantly impacting quality of life and imposing substantial financial burdens. Technology-delivered self-help interventions, including internet-based cognitive behavioral therapy (iCBT), have emerged as scalable and cost-effective solutions, particularly for underserved populations. However, fully self-administered interventions often struggle with limited effectiveness and low engagement. Recently, natural language processing (NLP) has been applied to enhance treatment effectiveness and adherence by improving user experience through personalized interactions. Despite promising applications, previous studies have primarily compared NLP-based interventions to passive controls or heterogeneous alternatives, making it difficult to isolate NLP's specific impact.

Objective

This exploratory study aimed to evaluate the effectiveness of NLP-supported iCBT in a fully self-administered setting by directly comparing it with a standard iCBT program (without NLP) and a waitlist control, using a structured exploratory randomized controlled trial.

Methods

We implemented a randomized, double-blind, parallel-group exploratory trial comparing AI-supported iCBT (with NLP advisory or empathetic feedback) to standard iCBT (without NLP feedback), using an otherwise identical program, alongside a waitlist control. Participants ($N = 1,187$) were recruited online and randomized if they met the inclusion criterion ($\text{PHQ-9} \geq 5$). The primary outcome was the proportion of participants with PHQ-9 scores ≥ 10 at Week 7 and Month 3. Secondary outcomes included changes in depressive and anxiety symptoms, functional impairment, adherence, and user satisfaction. Analyses followed the modified intention-to-treat (mITT) principle, and mixed-effects models for repeated measures (MMRM) were used, adjusting for baseline imbalances and stratification factors (age, gender, baseline PHQ-9).

Results

At Week 7, the standard iCBT group showed the largest short-term reduction in the proportion of participants with PHQ-9 scores ≥ 10 (21.4%), followed by AI-iCBT (41.4%) and waitlist (40.0%). However, by Month 3, AI-iCBT demonstrated a significant reduction to 19.7% ($p = 0.046$ vs control), while the iCBT group rebounded to 29.3% ($p = 0.413$ vs control), suggesting more sustained effects with AI-iCBT. Adherence trends indicated lower early participation in the AI-iCBT group but better retention after Week 4 ($p = 0.026$ for time-by-group interaction).

Conclusions

To our knowledge, this is the first randomized trial directly isolating the impact of NLP on depression in fully self-administered iCBT. Results suggest that while standard iCBT may provide faster short-term symptom relief, AI-enhanced iCBT offers more sustained improvement in depressive symptoms and may reduce the prevalence of moderate to severe depression over time. These findings highlight the potential of NLP to optimize digital mental health interventions, although further refinement and rigorous evaluation are warranted in future research.

Funding

The RIETI, Research Institute of Economy, Trade and Industry (a thinktank under the Ministry of Economy, Trade and Industry of the Government of Japan)

Introduction

Depression is a leading global cause of disability [1], significantly impairing quality of life [2, 3] and causing substantial financial burdens, including medical expenses [4] and productivity losses [5]. The rising prevalence, coupled with a shortage of healthcare resources, poses considerable strain for health systems and professionals in meeting the growing demand [6-9].

In response, technology-delivered self-help interventions have emerged as a promising solution for managing mental health difficulties. Most of these interventions are based on cognitive behavioral therapy (CBT) and are commonly referred to as internet-based CBT (iCBT) [10, 11]. These interventions guide users to independently manage their symptoms without human support, offering potential benefits such as cost reduction, alleviation of healthcare providers' burden, and improved access to mental health services, particularly in areas where such services are geographically, logistically, or institutionally inaccessible [9, 12]. While these interventions show promise, challenges remain in achieving treatment effects significant enough to promote recovery and in sustaining user engagement [13, 14].

Compared to blended formats (integrated into expert-led mental health care) or guided formats (supported by non-experts), which are more commonly implemented in clinical practice, fully independent self-help interventions often demonstrate limited effectiveness [15]. These interventions, delivered without external support, require further enhancement to maximize their potential and scalability in real-world settings. Recently, Natural Language Processing (NLP), an AI technology for understanding and generating human language [16, 17], has been applied to improve adherence and engagement through tailored interactions [18].

While systematic reviews highlight NLP's applications in mental health [18-22], its specific contributions to therapeutic effects and engagement have yet to be clearly demonstrated. This uncertainty stems from the use of heterogeneous control conditions in prior studies, such as waitlists, no intervention, or alternative modalities like treatment as usual, bibliotherapy, or conversational computer-based programs [22]. Consequently, it remains uncertain whether NLP provides distinct therapeutic benefits or merely functions as an active placebo. To address this, we implemented a structured randomized, parallel-group, exploratory clinical trial using a fully self-administered iCBT program in which conditions differed solely by the presence or absence of NLP feedback, enabling a double-blind assessment of its effects.

Methods

This study was a randomized, three-arm trial with partial double-blinding between the AI-iCBT and iCBT groups. The intervention arms included: AI-iCBT, a fully self-administered iCBT program incorporating NLP feedback; and iCBT, a fully self-administered program without NLP feedback. The waitlist (WL) group served as a control. Approval was granted by Hiramatsu Memorial Hospital's ethics committee in August 2015 (Reference number: 20150807), and the trial was registered (UMIN00001922), in accordance with the CONSORT 2010 Statement.

1) Study participants

Invitation emails were sent to all registered monitors of the research company (Nikkei Research Inc.) to recruit at least 900 participants. Interested individuals were instructed to complete the online Patient Health Questionnaire-9 (PHQ-9) [23, 24] for depression screening.

Participants were eligible if they met the following criteria: (1) aged 20–60 years at the start of the study, (2) access to the internet, (3) a baseline PHQ-9 score ≥ 5 , and (4) ability to understand Japanese and provide informed consent. Exclusion criteria included: (1) a medical condition precluding participation as determined by a physician, (2) concurrent participation in another CBT

program, (3) schizophrenia, (4) severe suicidal ideation, (5) dementia, or (6) a history of substance dependence (excluding smoking) in the past 12 months. Eligible participants were registered and randomly allocated to study groups. Just prior to the intervention, PHQ-9 scores were reassessed; individuals scoring <5 (i.e., in the normal range) were excluded from the efficacy analysis set (EAS) but remained in the overall study. Each participant received a 500-yen (approximately USD 4.50) cash voucher per assessment, provided at the final assessment point.

2) 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 program includes a self-guided cognitive restructuring □CR□ exercise, where users complete a seven-column thought record to address cognitive distortions. The NLP system processes user inputs, including facts, automatic thoughts, and feelings, referencing a corpus of past responses (Figure 1). It provides two types of automated feedback with phonation: empathetic messages delivered through an animated character whose expressions, such as smiling or showing sadness, are synchronized with the message content, and 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 (e.g., a feeling given instead of a thought) (Figure 2). In contrast, the non-AI iCBT program retained the same structure but provided only neutral, non-contextual responses, such as generic phrases like "Uh-huh" with neutral facial expressions. Both programs were otherwise identical in content, using the same validated iCBT package, which had previously demonstrated antidepressant effects [25]. This package consisted of six 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. The only difference between the two programs was the presence or absence 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 (9%) and advisory (4%) feedback functions (Figure 3).

3) Randomization and masking

The registered participants were randomly and concurrently assigned to either the AI-iCBT, iCBT, or WL groups according to a random sequence generated through computational methods by an independent third party, by stratified randomization based on age (≤ 40 vs. > 40), gender, and PHQ-9 score (≤ 9 vs. > 9). Participants assigned to the WL group were aware of their allocation (i.e., were non-blinded), whereas those in the AI-iCBT or iCBT groups were blinded and were only informed that they would be engaging in iCBT exercises using "the latest technology." This setup effectively ensured a double-blind condition between the AI-iCBT and iCBT groups.

4) Study procedures

E-mail reminders were automatically sent to participants twice a week during the 7-week exercise period. In each session, participants were initially required to view the online CBT psychoeducation e-learning and then conduct their allocated (AI-iCBT or iCBT) CR exercise at least once a week (six times or more in total). Those in the WL group did not undergo any exercise. All participants were also required to undergo assessments at baseline, post-exercise (Week 7), and at the end of follow-up (three months after the post-exercise assessment). The relevant information, including the attendant statuses or assessment results, was collected online.

5) Outcomes

The primary outcome was the proportion of participants with a PHQ-9 score ≥ 10 , which corresponds

to the threshold for major depression [23]. Secondary outcomes included mean scores on the PHQ-9 and the Quick Inventory of Depressive Symptomatology (QIDS-J) for depression [26, 27], the Generalized Anxiety Disorder-7 (GAD-7) for anxiety [28, 29], and the Sheehan Disability Scale (SDISS) for functional impairment [30, 31]. All instruments had been validated in Japanese and were assessed at three time points: baseline, Week 7 (post-intervention), and Month 3 (follow-up). In addition, user satisfaction was assessed using the Client Satisfaction Questionnaire-8 (CSQ-8) [32] at Week 7, but only among participants in the two intervention groups. Furthermore, the weekly exercise attendance rate was estimated as an additional index of engagement, counting each participant only once per week regardless of the number of sessions attended.

6) Statistical analyses

The sample size was estimated based on an assumed minimum effect size of 0.10 (Cohen's d) between the AI-iCBT and iCBT groups, due to the absence of directly comparable prior studies. A dropout rate of 50% was anticipated, reflecting patterns observed in previous similar research. The significance level was set at $\alpha = 0.05$ with 80% power. As this was an exploratory study focused on the contribution of AI, uncorrected nominal p -values were reported.

The primary analysis was conducted in the EAS following the modified intention-to-treat (mITT) principle (baseline PHQ-9 ≥ 5 and attendance of at least three sessions). Demographic and baseline characteristics were compared using one-way ANOVA or chi-square tests. The results revealed variability in the proportion of participants with PHQ-9 scores ≥ 10 at baseline: 67 (36.4%) in the AI-iCBT group, 70 (44.9%) in the iCBT group, and 143 (45.1%) in the waitlist group. Additionally, the mean baseline PHQ-9 scores were 8.8 (SD = 5.10) in the AI-iCBT group, 9.3 (SD = 4.90) in the iCBT group, and 10.3 (SD = 4.63) in the waitlist group, with statistically significant differences among groups ($p = 0.0021$), suggesting an imbalance in baseline severity. Since these differences were not anticipated, a post-hoc analysis was conducted controlling for baseline imbalances. The model included randomization stratification factors (age, gender, and baseline PHQ-9 score) and baseline outcome values as covariates. Outcomes were analyzed using mixed-effects models for repeated measures (MMRM), with intervention, time, and their interaction as fixed effects. A general covariance structure was applied, and results were reported as least squares means with 95% confidence intervals. Linear or logit link functions were used for continuous and binary outcomes, respectively. Missing data were handled within the MMRM framework. For participation rates, generalized MMRM with a logit link was used. An exploratory subgroup analysis divided participants into "mild" (PHQ-9 5–9) and "moderate-to-severe" (PHQ-9 ≥ 10) at baseline, reflecting the dual aims of preventing worsening in subthreshold cases and improving clinical depression. Outcomes were categorized accordingly at Week 7 and Month 3. A sensitivity analysis assessed dropout patterns based on session attendance. All analyses were performed using SPSS version 26.0 (IBM Japan).

8) Role of the funding source

The funder of the study had no role in study design, data collection, data analysis.

Result

1) Disposition of Subjects (Figure 4)

A total of 1,187 participants were eligible and randomly allocated to the AI-iCBT ($n=396$), iCBT ($n=397$), or WL ($n=394$) groups (ALL population). Participants who attended fewer than three sessions or had a PHQ-9 score < 5 at baseline (non-depressive) were excluded from the analyses, yielding 149 (37.6%) in AI-iCBT, 134 (33.8%) in iCBT, and 317 (80.5%) in the WL group,

respectively (EAS population). The number of participants who received post-intervention assessment at Week 7 was 131 in AI-iCBT, 122 in iCBT, and 269 in WL. At the follow-up, these numbers were 136, 122, and 260, respectively (Figure 4). After adjustment (Table 1), the mean PHQ-9 scores were 10.5 (SD = 4.51) for AI-iCBT, 10.5 (SD = 4.36) for iCBT, and 10.3 (SD = 4.36) for WL ($P = 0.932$ overall; $P = 0.939$ for AI-iCBT vs. iCBT), confirming that baseline severity of depressive symptoms was balanced across groups. Additionally, since medical history showed a significant baseline imbalance between groups ($p = 0.0383$), it was included as a covariate in the primary analysis to further account for potential confounding effects.

2) Efficacy Endpoint (Figure 5)

The primary analysis confirmed that the adjusted proportions of participants with PHQ-9 scores ≥ 10 at baseline remained comparable across groups: 48.6% in AI-iCBT, 50.5% in iCBT, and 47.7% in WL, with no significant differences between groups. A significant time-by-group interaction ($P = 0.008$) indicated that changes in depressive symptoms over time differed significantly among the three groups.

By Week 7, the proportions had reduced to 41.4% in the AI-iCBT group and 40.0% in the control group, though these reductions were not statistically significant. In contrast, the iCBT group demonstrated a substantial reduction to 21.4% ($P < 0.001$), which was significantly greater than the reduction in the control group ($P = 0.015$).

By the 3-month follow-up, the AI-iCBT group, which had shown limited improvement at Week 7, continued to show a significant reduction in the proportion of participants with PHQ-9 ≥ 10 to 19.7% ($P = 0.046$). A significant time-by-group interaction was also observed compared to the control group ($P = 0.028$), indicating a sustained and accelerated reduction in depressive symptoms. In contrast, the iCBT group experienced a rebound, with the proportion increasing to 29.3%, eliminating the significant advantage seen at Week 7 ($P = 0.413$). The control group also exhibited a significant reduction from baseline to Month 3 (from 47.7% to 34.1%, $P = 0.006$). However, this decrease was more gradual and less pronounced compared to both intervention groups. Notably, the rebound in the iCBT group by Month 3 highlights a potential limitation of this approach, in contrast to the sustained improvements seen in the AI-iCBT group. Thus, AI-iCBT appears to provide not only significant reductions in depressive symptoms over time but also more consistent long-term effects, as evidenced by the absence of a rebound.

3) Other Efficacy Outcomes

Regarding mean PHQ-9 scores (Figure 6-A), all groups showed significant within-group reductions by Week 7. Scores decreased to 8.89 in the AI-iCBT group ($P < 0.0001$), 8.48 in the iCBT group ($P < 0.0001$), and 9.52 in the control group ($P = 0.001$). Moreover, a significant time-by-group interaction was observed only for the iCBT group compared to the control group at Week 7 ($P = 0.019$), indicating that the iCBT group showed a greater reduction in depressive symptoms by this time point. At Month 3, the AI-iCBT group's score further decreased to 8.42, while the iCBT group's score remained stable at 8.49. The control group's score continued to decrease slightly, reaching 9.10. By this time point, the differences between the iCBT group and the control group were no longer statistically significant ($P = 0.107$). The overall time-by-group interaction was also not significant ($P = 0.128$).

Regarding QIDS-J scores, as indicated in Figure 6-B, similar to the PHQ-9 scores, by Week 7, only the iCBT group showed a significant reduction compared to the control group ($P = 0.0019$), while both the AI-iCBT and control groups showed slight and similar reductions. By Month 3, the AI-iCBT group's scores continued to decrease in an almost linear fashion, while the control group's decrease became more gradual. Meanwhile, the iCBT group experienced an upward rebound after reaching a low point at Week 7. By Month 3, neither the AI-iCBT nor the iCBT groups showed a statistically significant difference compared to the control group. The overall time-by-group interaction was also not significant ($P = 0.078$).

Regarding GAD-7 Scores, as shown in Figure 6-C, both the AI-iCBT and iCBT groups showed similar reductions by Week 7. Both groups demonstrated significant within-group reductions compared to baseline ($P < 0.0001$ for both), but neither group exhibited a statistically significant difference from the control group. By Month 3, both AI-iCBT and iCBT groups continued to show modest reductions, but neither group showed a statistically significant difference compared to the control group. Unlike the results for PHQ-9 and QIDS-J, where significant differences were observed, the reductions in GAD-7 scores were smaller and comparable to the control group, with no significant time-by-group interaction ($P = 0.704$).

For SDS, by Month 3, both the AI-iCBT and iCBT groups showed significant within-group reductions ($P = 0.01$), but neither group showed a statistically significant difference compared to the control group at any time point. Regarding CSQ-8, at Week 7, the AI-iCBT and iCBT groups showed no significant differences in overall CSQ-8 scores or any individual items. The total score for the AI-iCBT group was 21.11, and for the iCBT group, it was 20.68, with a mean difference of 0.4 points ($P = 0.417$). All P-values for individual items were above the significance threshold, indicating similar levels of satisfaction between the two groups.

4) Exercise Participation Rates

Both the AI-iCBT and iCBT groups experienced gradual declines in participation rates over the intervention period, fluctuating between 80% and 95%, indicating a relatively high level of engagement was sustained. However, distinct trajectories emerged between the groups (Figure 7). Initially, iCBT demonstrated higher participation rates up to Week 3, but by Week 4, AI-iCBT surpassed iCBT, maintaining this advantage in the following weeks. While the overall difference between the groups was not statistically significant ($P = 0.216$), a significant group-by-time interaction ($P = 0.026$) suggested that AI-iCBT followed a notably different participation trajectory, with superior retention of engagement from Week 4 onward.

5) Sub-group Analysis

Changes in the “mild” group from baseline to the final assessment point at Week 7 or Month 3, classified into exacerbated (PHQ-9 ≥ 10), “stable (PHQ-9 5 to 9),” and “improved (PHQ-9 < 5),” were shown for each intervention group in Figure 8-A. Similarly, changes in the “moderate-to-severe” group were shown, classified into “not improved (PHQ-9 ≥ 10)” and “improved (PHQ-9 < 10)” (Figure 8-B).

5) Sensitivity Analysis

Between those who attended three or more sessions and those who attended fewer than three, significant differences were observed in gender, age, and employment status, with the lower-attendance group having more males ($P < 0.0001$), older individuals ($P = 0.0012$), and employed

participants ($P=0.0157$). In contrast, no significant differences were found in baseline PHQ-9, marital status, educational background, medical history, or mental health history.

Discussion

To the best of our knowledge, this study is considered unique in directly verifying the enhanced antidepressant effects of AI (NLP) functionality on a fully self-administered cognitive restructuring program (iCBT), with conditions differing solely in the presence or absence of NLP feedback, within the framework of a randomized, double-blind, controlled trial design. Notably, only iCBT with NLP feedback demonstrated a significant reduction in the proportion of participants with major depression levels of symptoms compared to the control group at the 3-month follow-up. Furthermore, we observed that the presence or absence of NLP feedback alone can significantly influence participation rates in iCBT exercises. While NLP-based interventions have advanced significantly, their specific contributions to depression or anxiety treatment have still remained unclear due to the use of heterogeneous control conditions in prior studies. This study, made possible by its more robust design, provides new insights into how AI can enhance antidepressive interventions and prevention, offering valuable perspectives in mental health care.

While it might be intuitive to assume that the addition of NLP feedback would simply enhance the antidepressant effects of iCBT, the addition actually appear to have led to complex interactions, including instances where it may have temporally attenuated these effects. During the 7-week intervention, only iCBT without NLP feedback showed a significant antidepressant effect in the continuous scores of PHQ-9 and QIDS-J compared to the control group. However, in the follow-up period, its effects plateaued (PHQ-9) or rebounded after reaching a floor (QIDS-J), ultimately showing no significant differences from the control group. By contrast, while the binary PHQ-9 ≥ 10 proportion mirrored the continuous scores during the intervention period, only AI-iCBT demonstrated a significant antidepressant effect against the control group at the 3-month follow-up—a benefit absent in conventional iCBT. This divergence between binary and continuous measures suggests the potential involvement of a more intricate underlying mechanism.

A sub-group analysis revealed two distinct patterns in the effects of AI-iCBT and iCBT, which appeared to vary by baseline severity. First, non-AI iCBT seemed to show more pronounced antidepressant effects for individuals with mild depression at baseline, whereas AI-iCBT appeared to be more beneficial for those with moderate-to-severe depression at baseline, pointing to potential differences in their optimal target zones. Second, AI-iCBT appeared to have an advantage over non-AI iCBT in potentially preventing the future exacerbation of mild depression, indicating its possible

role in mitigating the risk of worsening symptoms over time. Taken together, these findings suggest that non-AI iCBT without NLP feedback may offer short-term relief for subthreshold depression, while NLP-enhanced iCBT shows potential for treating and preventing major depression.

These gaps, purely attributable to NLP feedback, raise questions about its contribution. NLP feedback includes two components: empathetic feedback, aimed at relieving emotional distress and fostering support [33, 34], and advisory feedback, designed to enhance the fidelity of CR exercises for greater antidepressant effects [35]. Together, these elements were expected to complement each other, addressing both emotional and procedural needs to strengthen the intervention's impact. Additionally, both functions were hypothesized to mitigate common challenges associated with fully self-administered iCBT, such as users' uncertainties or lack of confidence about the correctness of their inputs, by providing empathy and guidance. The interactive nature of the feedback was also believed to transform a potentially monotonous task into a more engaging process, increasing satisfaction and subsequently boosting engagement and adherence—factors anticipated to contribute to the differences in therapeutic effects. However, concerns arose during the intervention period that advisory feedback, in particular, might have been perceived as overly critical. This perception could have potentially demotivated users and temporarily diminished short-term antidepressant effects.

In this study, it was practically difficult to disentangle the individual contributions of advisory and empathetic feedback due to their apparent simultaneous operation in the post-hoc analysis. However, the lack of significant differences between the groups in the several observed results suggests that our initially hypothesized factors may not adequately explain their impact. Specifically, NLP feedback did not appear to contribute to anxiety reduction, as indicated by GAD-7 scores, nor did it enhance satisfaction, as shown by CSQ-8 scores. Furthermore, its influence on participation motivation, as reflected in exercise participation rates, seemed limited. Regarding antidepressant effects during the intervention period, iCBT without NLP feedback unexpectedly demonstrated greater efficacy.

On the other hand, a significant group-by-time interaction in participation rates was observed. While AI-iCBT showed lower participation rates early in the intervention, a favourable trend emerged in the latter half. In iCBT, a correlation between greater antidepressant effects and lower dropout rates has been reported [36]. If AI-iCBT had demonstrated superior antidepressant effects during the intervention period, this correlation might explain the observed dynamics. However, the reality was the opposite, as AI-iCBT underperformed compared to iCBT in terms of antidepressant effects at

least during the intervention period. This suggests that the dynamics observed in this study may stem from mechanisms distinct from the commonly recognized relationship between antidepressant effects and adherence.

Notably, these unknown mechanisms might have been factors that specifically contributed to the effects observed in current and future cases of major depression. Although the precise nature of these mechanisms remains unclear, latent factors mediated by inter-species communication between the AI and participants could have played a role in these results. In face-to-face psychotherapy, approaches such as interpersonal psychotherapy, as well as family interventions and family psychoeducation—which all more explicitly focus on relationships—have often been reported to demonstrate delayed effects[37-42]. While especially interpersonal psychotherapy may be less immediately effective than CBT during the intervention period, some studies suggest that these relationally focused approaches may even surpass CBT in the long term, particularly after the intervention has ended[37-39]. This pattern of delayed yet potentially superior impact might, in some form, be mirrored in the way AI-based feedback interacts with participants in digital interventions.

The phenomena observed in this study might have remained unclear under heterogeneous control conditions but were clarified through the more structured RCT design employed here. Despite the advancements in AI communication, including generative AI, logistical challenges and high development costs often hinder the creation of suitable control programs. This limitation has left the full potential of AI communication, including its antidepressant and anxiolytic effects, underexplored. Addressing these issues is essential to unlocking AI's broader applications in psychosocial interventions or further educational and rehabilitation programs.

This study has several limitations. First, most participants were tech-savvy research panel members, with 12% being active mental health service users. While this aligns with the average mental health service utilization rate among the Japanese population, caution is needed when generalizing these findings to broader populations. Second, a high dropout rate was observed, with fewer than 50% of EAS participants completing at least three sessions. This is consistent with previous large-scale, fully self-administered intervention studies, which often face challenges in sustaining engagement, especially when including cases with missing post-treatment data[43]. The sensitivity analysis revealed that lower attendance was associated with being male, older, and employed, indicating that time constraints, rather than health-related factors, may have been a primary barrier to participation, particularly for busier working men[44]. In addition to addressing missing data through mixed-

effects models, we adjusted for group differences using covariates such as age, gender, and baseline PHQ-9 scores, which were stratified during randomization. Still, careful interpretation remains necessary. Third, further investigation is needed to fully understand the mechanisms underlying the intervention's effects. Specifically, it was not possible to isolate the individual therapeutic contributions of advisory and empathetic feedback, as both operated concurrently. Additionally, due to technical issues, session-by-session data on depressive symptoms (ODSIS) and anxiety (OASIS) were lost, limiting the ability to conduct more granular analyses. Lastly, the overall effect size was small, potentially due to a ceiling effect, as participants with milder depression had limited room for improvement.

Despite these challenges, this study provides valuable insights into the potential of AI-enhanced self-help digital interventions, particularly in relation to long-term outcomes post-intervention, as well as subgroup-specific therapeutic responsiveness and 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.

Acknowledgments

We are grateful to Dr. Yutaka Ono (Center for the Development of Cognitive Behavior Therapy Training, Tokyo), Professor Takashi Watanabe (Teikyo Heisei University, Tokyo), Mr. Kenichiro Tsumura (T Quest, Chiba), Dr. Sosei Yamaguchi (National Center of Neurology and Psychiatry, Tokyo), and all members of the project teams of NEC Solution Innovators, Ltd. (Tokyo) for allowing us to unrestrictedly utilize AI-iCBT for the study.

Declaration of interests

The authors have no COI to disclose.

Funding

The Research Institute of Economy, Trade and Industry (A thinktank under the Ministry of Economy, Trade and Industry of the Government of Japan).

Author 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. HY contributed to secondary analysis and writing—review and editing.

NW contributed to conceptualization, investigation, methodology, supervision, validation, and writing—review and editing.

Data sharing statement

The data used in the present study belong to the RIETI and can be obtained from the institute upon reasonable request.

Preprint
JMIR Publications

Table 1. Subjects’ demographic and medical condition characteristics (Efficacy Analysis Set)

Characteristic		Total	AI-iCBT group	iCBT group	Control group	P value-1 (ANOVA)	P value-2 (chi-square test)
	N	600	149	134	317		
Gender	Male	318 (53.0%)	72 (48.3%)	65 (48.5%)	181 (57.1%)		0.1038
	Female	282 (47.0%)	77 (51.7%)	69 (51.5%)	136 (42.9%)		
Age	Mean (SD)	42.5 (9.87)	42.1 (9.49)	41.2 (9.62)	43.2 (10.10)	0.1120	
(years)	Median	43	43	41	44		
	Min, Max	20, 60	21, 60	20, 60	20, 60		
Marital status	Married	320 (53.3%)	74 (49.7%)	76 (56.7%)	170 (53.6%)		0.2678
	Divorced	41 (6.8%)	10 (6.7%)	4 (3.0%)	27 (8.5%)		
	Bereaved	4 (0.7%)	1 (0.7%)	2 (1.5%)	1 (0.3%)		
	Single	235 (39.2%)	64 (43.0%)	52 (38.8%)	119 (37.5%)		
Educational background	Junior high school	3 (0.5%)	1 (0.7%)	0 (0.0%)	2 (0.6%)		0.3868
	High school	115 (19.2%)	38 (25.5%)	22 (16.4%)	55 (17.4%)		
	Junior college, technical college,	115	27 (18.1%)	28 (20.9%)	60 (18.9%)		

	vocational school	(19.2%)					
	University, post-graduate	367 (61.2%)	83 (55.7%)	84 (62.7%)	200 (63.1%)		
Employment status	Working	472 (78.7%)	106 (71.1%)	106 (79.1%)	260 (82.0%)		0.0762
	Unemployed (looking for work)	48 (8.0%)	16 (10.7%)	13 (9.7%)	19 (6.0%)		
	Unemployed (not looking for work)	80 (13.3%)	27 (18.1%)	15 (11.2%)	38 (12.0%)		
Medical history	No relevant history	460 (76.7%)	119 (79.9%)	100 (74.6%)	241 (76.0%)		0.0383
	Ambulant	134 (22.3%)	26 (17.4%)	32 (23.9%)	76 (24.0%)		
	Hospitalized	6 (1.0%)	4 (2.7%)	2 (1.5%)	0 (0.0%)		
Mental history	In treatment	74 (12.3%)	21 (14.1%)	16 (11.9%)	37 (11.7%)		0.9158
	Treated	109 (18.2%)	29 (19.5%)	24 (17.9%)	56 (17.7%)		
	No relevant history	417 (69.5%)	99 (66.4%)	94 (70.1%)	224 (70.7%)		
Mean baseline scale score	PHQ-9≥10	284 (47.3%)	70 (47.0%)	71 (53.0%)	143 (45.1%)		0.3084
						Whole	AI-iCBT vs iCBT
	PHQ-9	10.4 (4.54)	10.5 (4.51)	10.5 (4.36)	10.3 (4.63)	0.9322	0.9390#
	QIDS-J	10.0 (4.40)	10.2 (4.32)	10.1 (4.31)	9.9 (4.48)	0.7578	0.8468#
	GAD-7	7.2 (4.33)	6.9 (4.04)	7.2 (4.29)	7.3 (4.48)	0.6697	0.5330#

Adjusted by gender, age, baseline PHQ-9 score

#: pairwise comparison of group AI-iCBT and iCBT

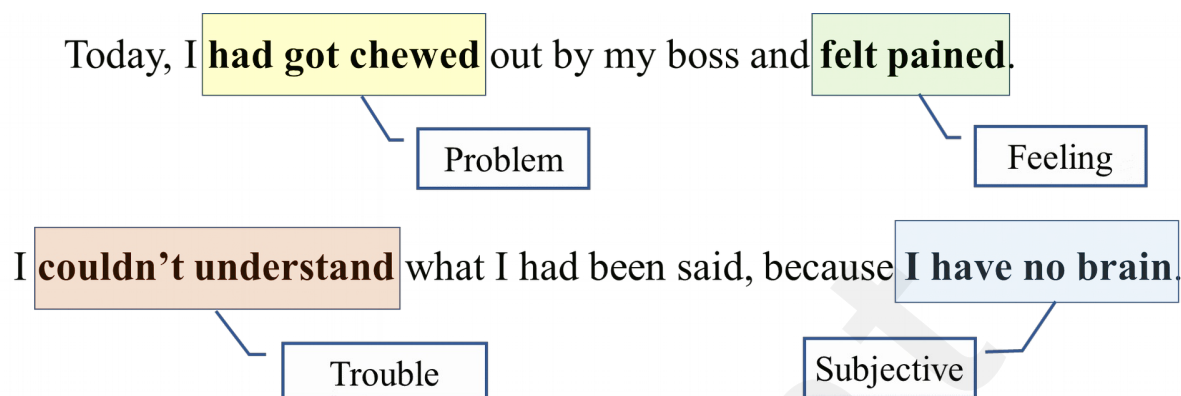
Figures

Figure 1: Example of an expression extracted from the corpus of NLP-driven problem statement
(original text in Japanese)

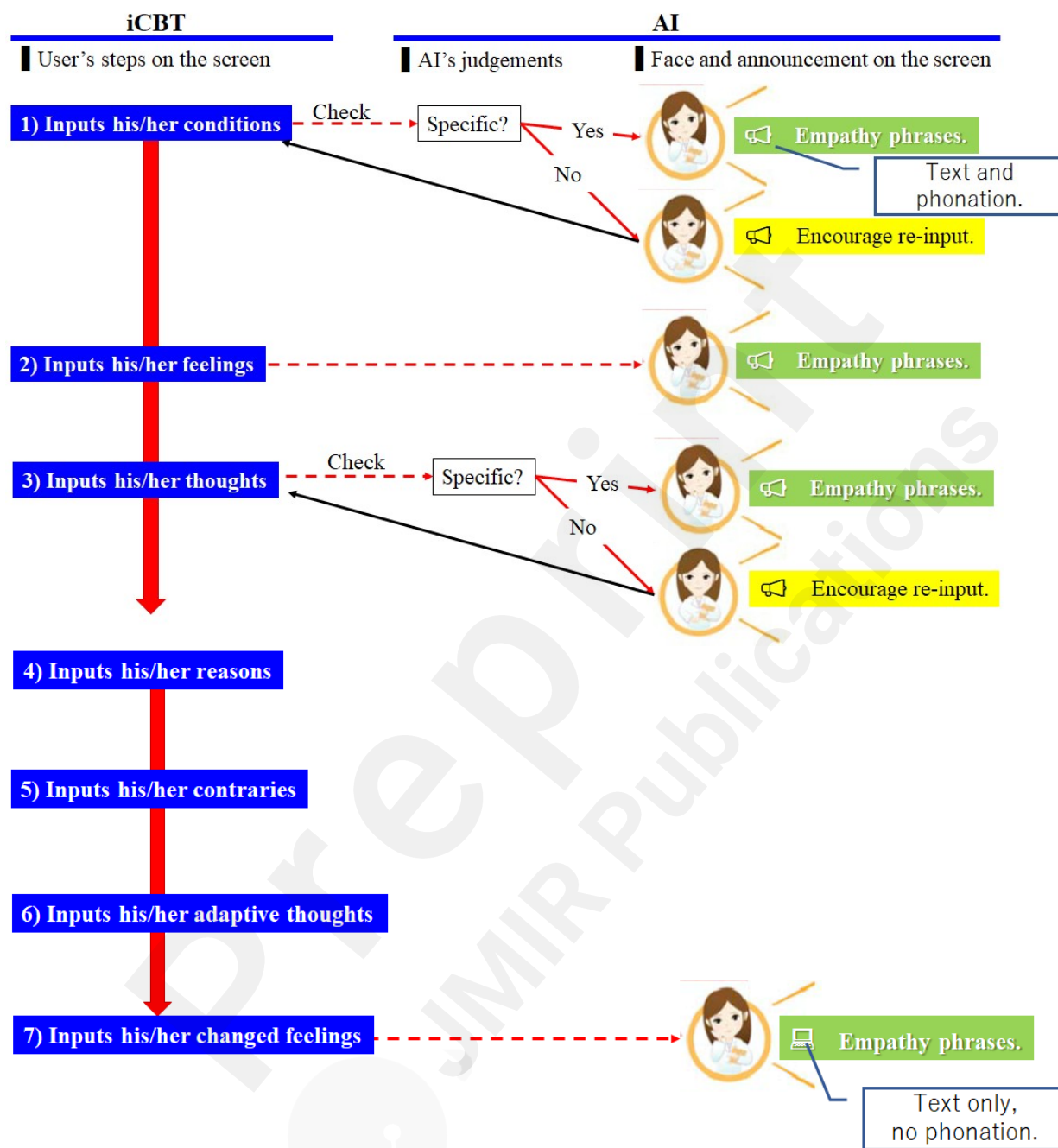


Figure 2: Procedure of AI-led internet-based cognitive behavioral therapy (AI-iCBT)

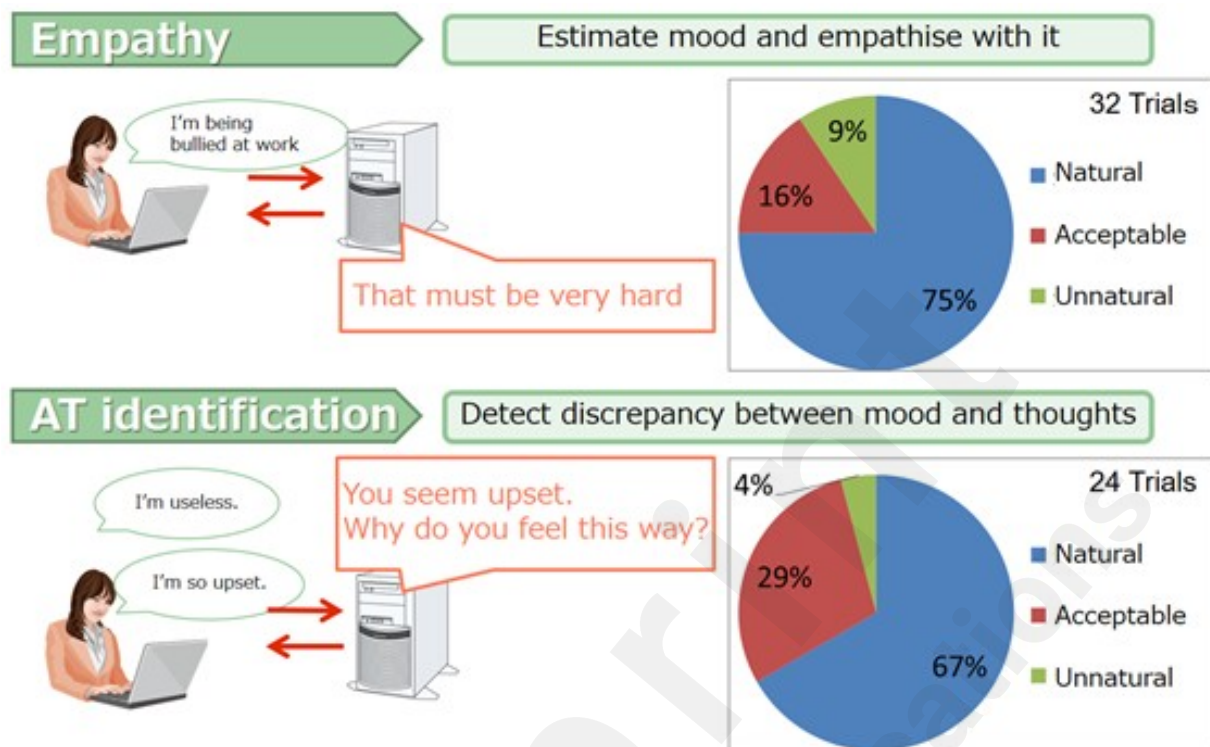


Figure 3: Qualitative preliminary analysis of talking therapy using the NPL-driven dialogue engine

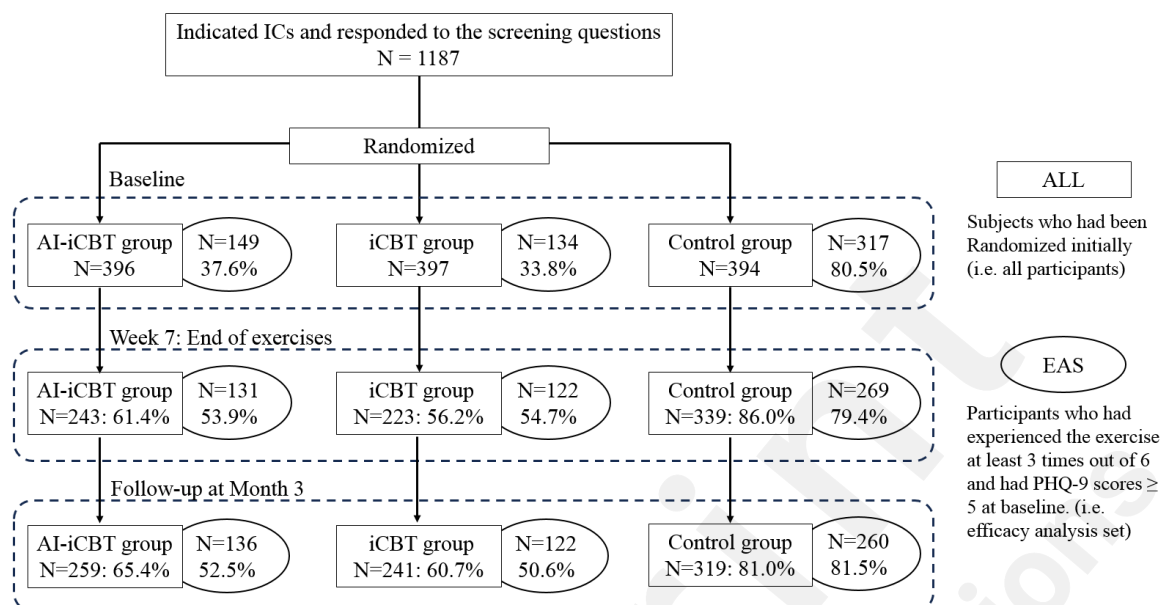


Figure 4. Subjects' disposition (for all and adherence participants)

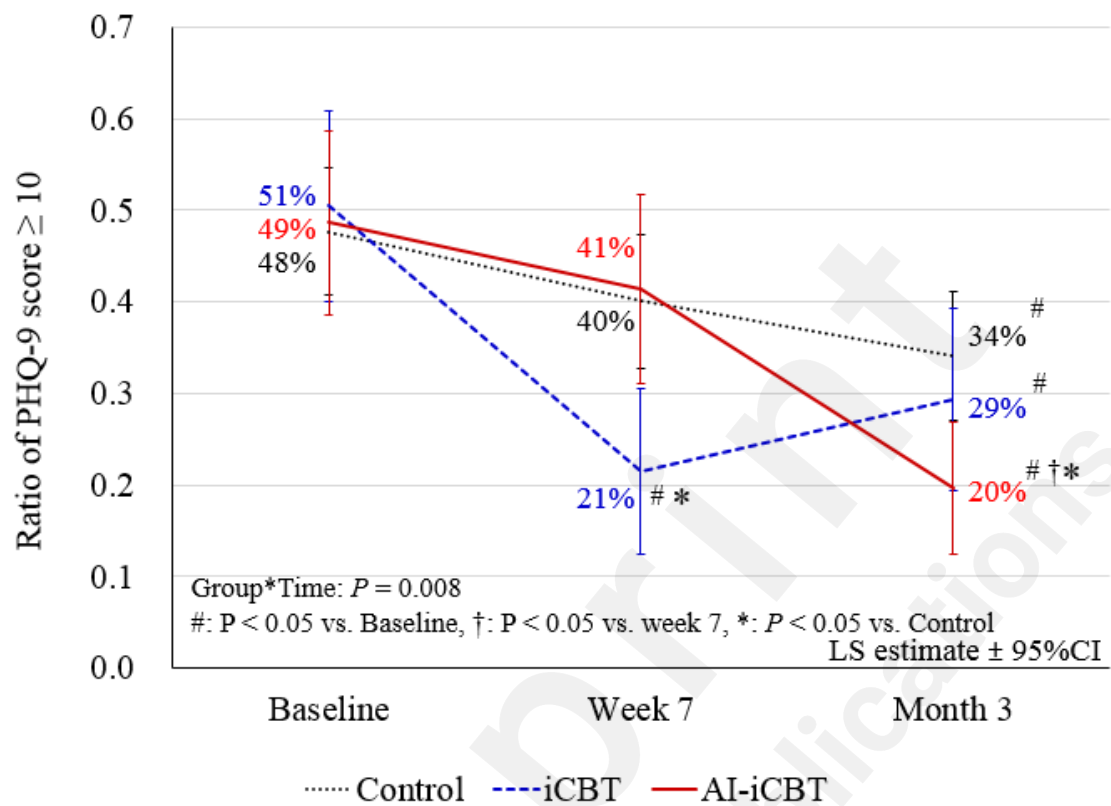


Figure 5. Percentages of subjects with a PHQ-9 score ≥ 10 at each evaluation point

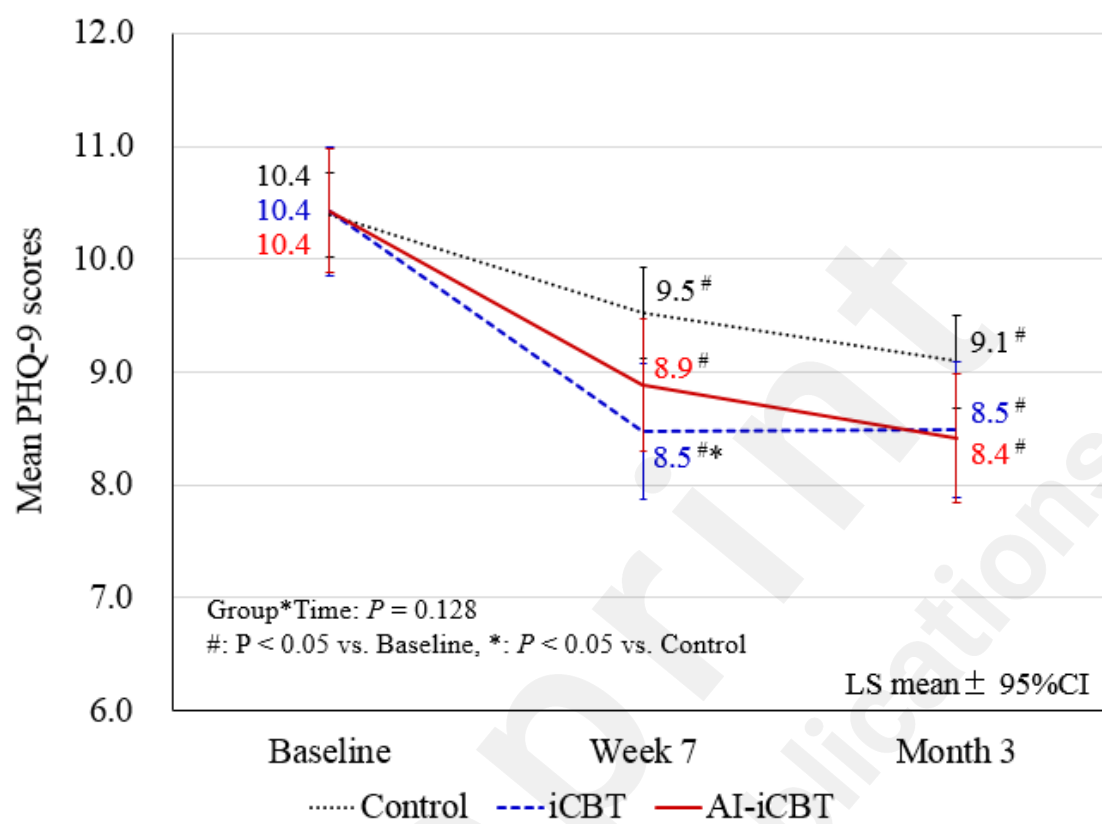


Figure 6-A. Changes in mean PHQ-9 scores

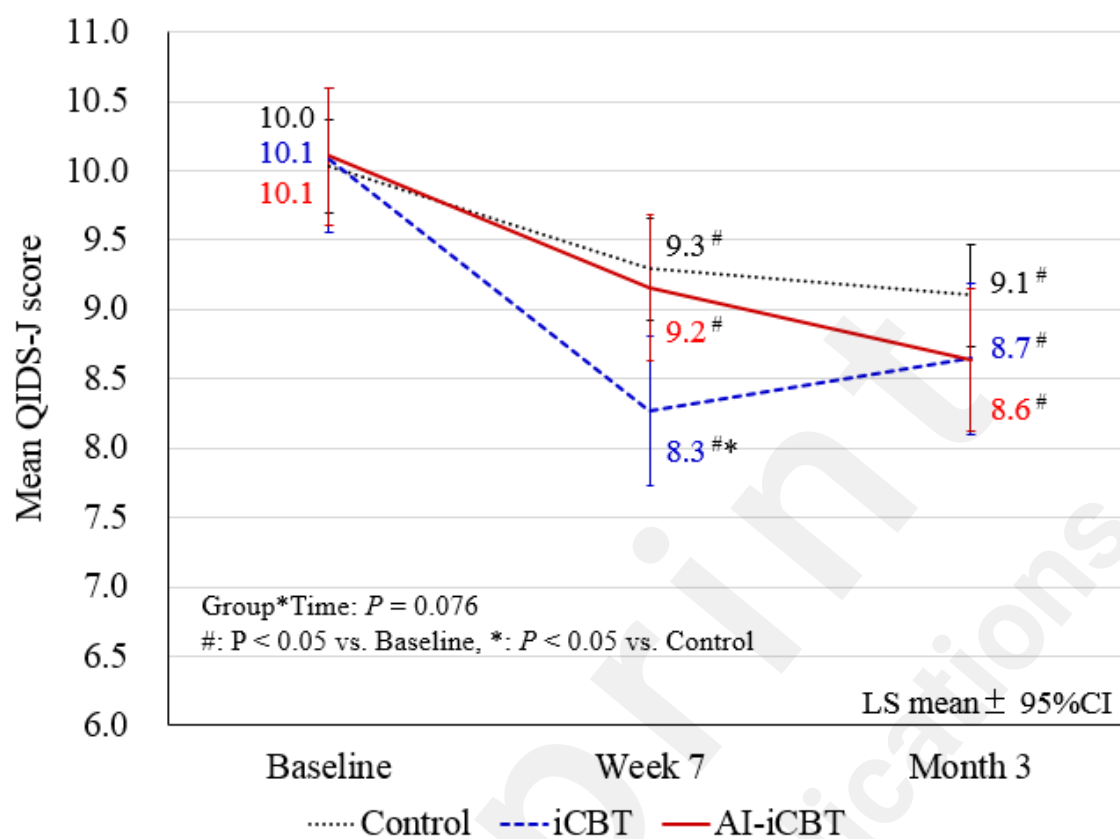


Figure 6-B. Changes in mean QIDS-J scores

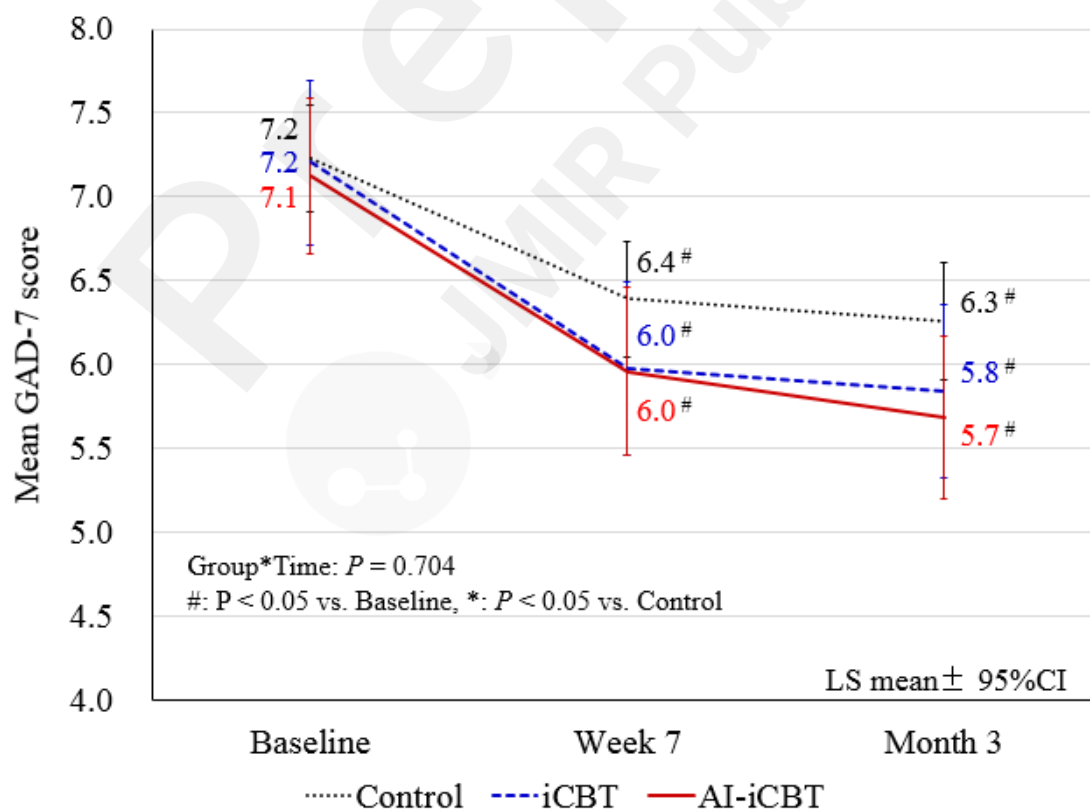
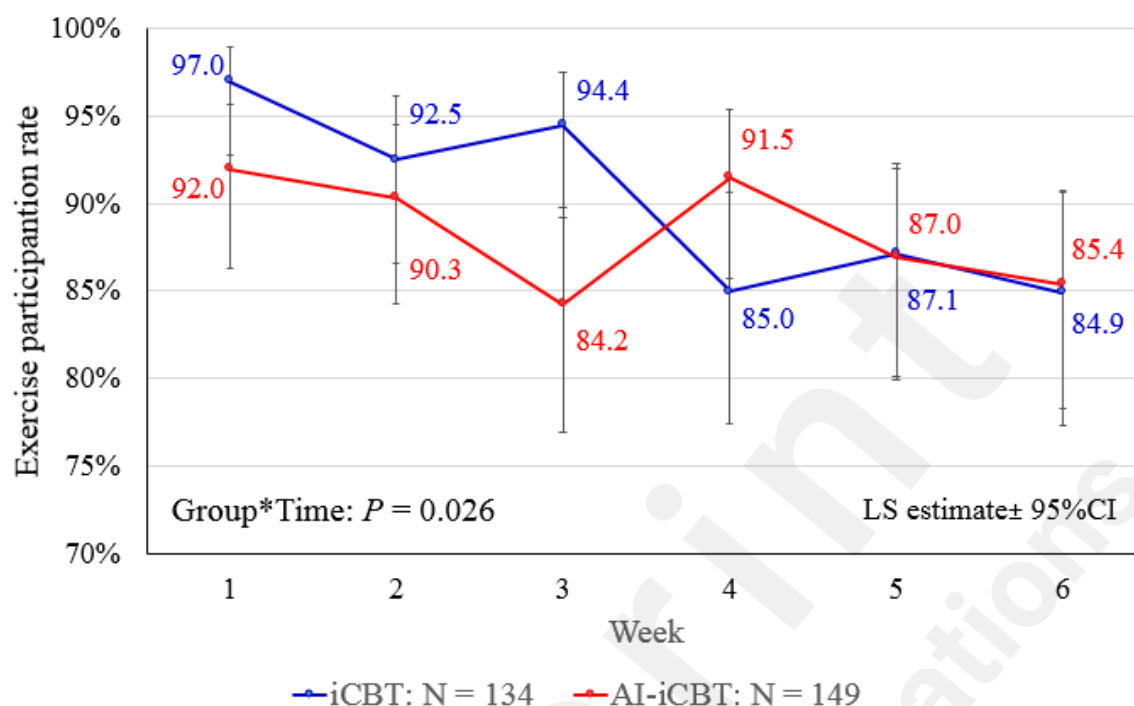
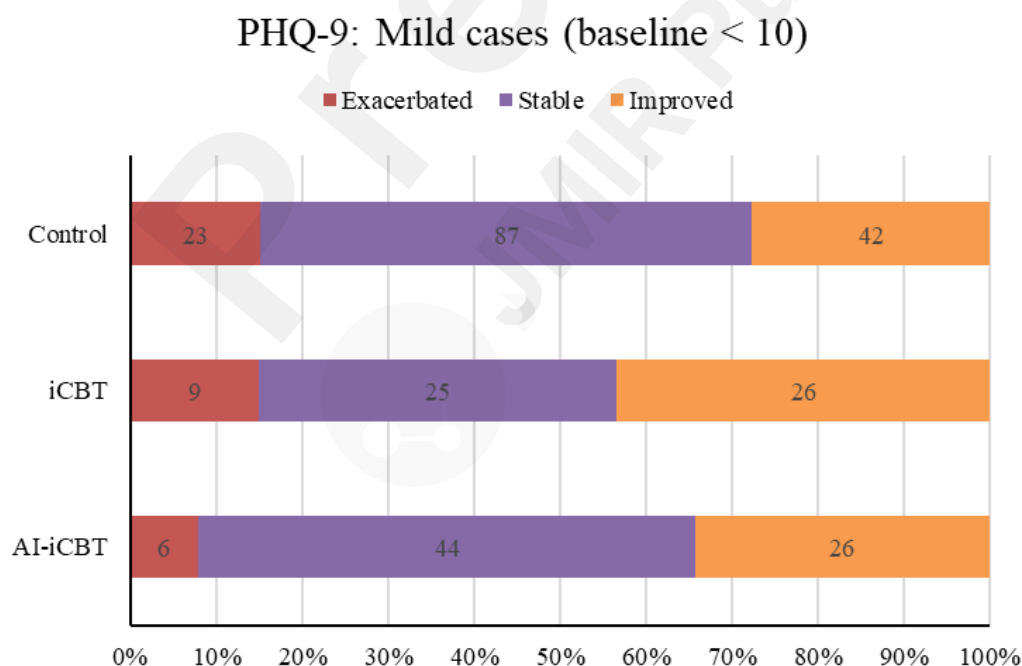


Figure 6-C. Changes in mean GAD-7 scores**Figure 7.** Time courses of exercise participation rates from Week 1 to 6 in the AI-iCBT and iCBT groups**Figure 8-A.** Final PHQ-9 status in participants with mild depression at baseline (PHQ-9 < 10)

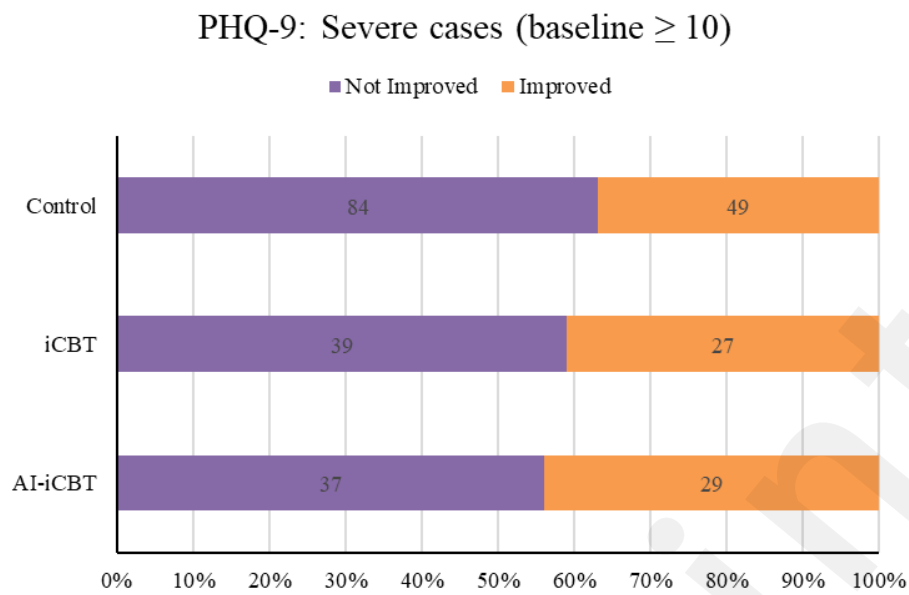


Figure 8-B. Final PHQ-9 status in participants with moderate-to-severe depression at baseline (PHQ-9 ≥ 10)

Preprint
JMIR Publications

Figure legends

Figure 1. Example of an expression extracted from the NLP-driven CBT problem statement corpus

A corpus of conflict-related texts from 28,718 past CBT cases was compiled, containing statements of problems, emotions, and subjective expressions. These data were used to generate response messages (voice prompts) within the dialogue engine that underpins the NLP system used in AI-iCBT.

Figure 2. Procedure of AI-led internet-based cognitive behavioral therapy (AI-iCBT)

Participants completed a structured 7-step exercise based on the standard cognitive restructuring thought record: (1) Situation, (2) Feeling, (3) Automatic thoughts, (4) Supporting evidence, (5) Contradictory evidence, (6) Alternative thought, and (7) Re-rated feeling. The NLP system was designed to function as part of the therapeutic process by providing real-time voice prompts and on-screen guidance with context-sensitive facial animations when input was judged to be vague or inappropriate, and by displaying empathetic feedback—also accompanied by facial animations—when the emotion field was filled in.

Figure 3. Qualitative preliminary analysis of talking therapy using the NLP-driven dialogue engine

A preliminary qualitative study was conducted to evaluate interactive support features of the NLP-driven dialogue engine. Discomfort was reported by only 9% (3/32) of participants for the empathy function, which estimated users' emotions, and by 4% (1/24) for the advisory function, which identified discrepancies between users' feelings and thoughts. These results suggest that the interactive support features were generally well accepted.

Figure 4. Subject disposition and inclusion in the efficacy analysis set (EAS)

This diagram illustrates the flow of participants from initial registration to inclusion in the efficacy analysis set. Among 1,187 registered participants, 396 were assigned to AI-iCBT, 397 to iCBT, and 394 to the waitlist control group. The efficacy analysis set (EAS) was defined as participants who had baseline PHQ-9 scores ≥ 5 and completed at least three of the six weekly sessions. As a result, 149 participants (37.6%) in the AI-iCBT group, 134 (33.8%) in the iCBT group, and 317 (80.5%) in the control group were included in the EAS. While the proportion of participants who met the adherence criteria was numerically higher in the AI-iCBT group (46.5%) than in the iCBT group (39.3%), this post hoc difference reached statistical significance ($P = 0.0446$, Fisher's exact test) and should be interpreted exploratorily.

Figure 5. Proportion of participants with PHQ-9 scores ≥ 10 at each time point (primary endpoint).

The proportion of participants with PHQ-9 ≥ 10 was comparable among groups at baseline ($P = 0.150$). At Week 7, only the iCBT group showed a significant reduction compared to the control group ($P = 0.015$), but this effect was not maintained at Month 3. In contrast, the AI-iCBT group exhibited a sustained reduction and showed a significantly lower proportion than the control group at Month 3 ($P = 0.046$), suggesting more enduring antidepressant effects.

Figure 6-A. Changes in mean PHQ-9 scores over time

Mean PHQ-9 scores decreased over time in all groups, including the control group. At Week 7, only the iCBT group showed a significant reduction compared to the control group, but

this effect was not sustained at Month 3. In contrast, the AI-iCBT group exhibited a continued decrease through Month 3, although the difference between the three groups at this time point was not statistically significant. These trends suggest that while AI-iCBT may have delayed initial improvement, it potentially contributed to more stable symptom reduction over time.

Figure 6-B. Change in mean QIDS-J scores over time

Mean QIDS-J scores decreased in all groups during the intervention period. At Week 7, only the iCBT group showed a significant reduction compared to the control group, but this improvement was not sustained, as scores rebounded by Month 3. In contrast, the AI-iCBT group showed a steady linear decline throughout the entire period. Although the differences among groups were not statistically significant at Month 3, these patterns suggest that AI-iCBT may have contributed to more consistent symptom improvement, while iCBT's short-term benefits diminished over time.

Figure 6-C. Change in mean GAD-7 scores over time

Mean GAD-7 scores declined modestly in all groups. Both AI-iCBT and iCBT groups showed significant within-group improvements by Week 7, but there were no statistically significant differences compared to the control group at any time point. Unlike PHQ-9 and QIDS-J scores, GAD-7 changes were smaller and similar across all groups, suggesting that the intervention had limited added benefit for anxiety symptoms.

Figure 7. Time courses of exercise participation rates from Week 1 to 6 in the AI-iCBT and iCBT groups

Participation rates were high (80–95%) across both intervention groups throughout the 7-week period. While iCBT showed higher initial engagement, AI-iCBT demonstrated improved retention from Week 4 onward. Although overall group differences were not statistically significant ($P = 0.216$), a significant group-by-time interaction ($P = 0.026$) suggests divergent patterns in participation trends.

Figure 8-A. Final PHQ-9 status in participants with mild depression at baseline (PHQ-9 < 10)

This figure shows the final PHQ-9 status (at Week 7 or Month 3) for participants who were classified as having mild depression at baseline. While there was no statistically significant difference in the proportion of participants with symptom exacerbation (PHQ-9 ≥ 10) among the groups ($P = 0.1075$), the AI-iCBT group exhibited the lowest exacerbation rate (8.0%) compared to iCBT (15.0%) and control (15.1%) groups. On the other hand, the iCBT group demonstrated the highest improvement rate (43.3%; defined as PHQ-9 < 5), suggesting better short-term symptom relief. These results indicate that AI-iCBT may be more effective in preventing symptom deterioration, whereas iCBT may lead to more rapid remission in mild depression.

Figure 8-B. Final PHQ-9 status in participants with moderate-to-severe depression at baseline (PHQ-9 ≥ 10)

This figure summarizes the outcomes of participants with moderate-to-severe depression at baseline. Although no statistically significant group differences were observed in the improvement rate (PHQ-9 < 10) ($P = 0.7555$), the AI-iCBT group had the highest proportion of improvement (45.3%), numerically outperforming both the iCBT and control groups. This suggests that AI-iCBT may be particularly beneficial in supporting symptom improvement among individuals with higher baseline severity.

References

1. James, S.L., et al., *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*. The Lancet, 2018. **392**(10159): p. 1789-1858.
2. Arias, D., S. Saxena, and S. Verguet, *Quantifying the global burden of mental disorders and their economic value*. EClinicalMedicine, 2022. **54**(101675).
3. Hohls, J.K., et al., *Anxiety, Depression and Quality of Life—A Systematic Review of Evidence from Longitudinal Observational Studies*. International Journal of Environmental Research and Public Health, 2021. **18**(22): p. 12022.
4. König, H., H.H. König, and A. Konnopka, *The excess costs of depression: a systematic review and meta-analysis*. Epidemiol Psychiatr Sci, 2019. **5**(29).
5. World Health Organization. *Mental health at work*. 2022 [cited 2024 Nov 25th]; Available from: <https://www.who.int/news-room/fact-sheets/detail/mental-health-at-work>.
6. Cunningham, P.J., *Beyond parity: primary care physicians' perspectives on access to mental health care*. Health Aff, 2009. **28**(3): p. 14.
7. Wang, P.S., 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): p. 841-50.
8. Wainberg, M.L., F.G. Lu, and M.B. Riba, *Global Mental Health*. Acad Psychiatry, 2016. **40**(4): p. 647-9.
9. Edge, D., et al., *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): p. 22.
10. World Health Organization. *Psychological interventions implementation manual: integrating evidence-based psychological interventions into existing services*. . 2024 [cited 2024 Nov 25th]; Available from: <https://www.who.int/publications/i/item/9789240087149>.
11. World Health Organization. *Consolidated telemedicine implementation guide*. : ; . Available from: <https://www.who.int/publications-detail-redirect/9789240059184> [accessed Jun 7, 2024]. 2022.
12. Saad, A., et al., *Self-directed Technology-Based Therapeutic Methods for Adult Patients Receiving Mental Health Services: Systematic Review*. JMIR Ment Health, 2021. **8**(11): p. 27404.
13. Bauer, A.M., et al., *Acceptability of mHealth augmentation of Collaborative Care: A mixed methods pilot study*. Gen Hosp Psychiatry, 2018. **51**: p. 22-29.
14. Weisel, K.K., et al., *Standalone smartphone apps for mental health-a systematic review and meta-analysis*. NPJ Digit Med, 2019. **2**(118): p. 019-0188.
15. Pineda, B.S., et al., *Updated taxonomy of digital mental health interventions: a conceptual framework*. Mhealth, 2023. **9**(28): p. 23-6.
16. Dwyer, D.B., P. Falkai, and N. Koutsouleris, *Machine Learning Approaches for Clinical Psychology and Psychiatry*. Annu Rev Clin Psychol, 2018. **14**: p. 91-118.
17. Nie, J., et al., *LLM-based Conversational AI Therapist for Daily Functioning Screening and Psychotherapeutic Intervention via Everyday Smart Devices*. CoRR, 2024. **abs/2403.10779**.
18. Malgaroli, M., et al., *Natural language processing for mental health interventions: a systematic review and research framework*. Transl Psychiatry. 2023 Oct 6;13(1):309. doi: 10.1038/s41398-023-02592-2.
19. Le Glaz, A., et al., *Machine Learning and Natural Language Processing in Mental Health: Systematic Review*. J Med Internet Res, 2021. **23**(5): p. 15708.
20. Zhang, T., et al., *Natural language processing applied to mental illness detection: a narrative review*. npj Digital Medicine, 2022. **5**(1): p. 46.
21. Choudhury, M.D., S.R. Pendse, and N. Kumar, *Benefits and Harms of Large Language Models in Digital Mental Health*. ArXiv, 2023. **abs/2311.14693**.
22. Villarreal-Zegarra, D., 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**: p. e59560.
23. Kroenke, K., R.L. Spitzer, and J.B. Williams, *The PHQ-9: validity of a brief depression severity measure*. J Gen Intern Med, 2001. **16**(9): p. 606-13.
24. Muramatsu, K., *Japanese version of Patient Health Questionnaire (PHQ-9, PHQ-15) and eneralized*

- Anxiety Disorder -7 -up to date- (In Japaese) The Japanese journal of clinical psychology, 2014. 7: p. 35-39.
25. So, M., Sekizawa, Y., Yamaguchi, Y., A Randomised Controlled Trial Investigating the Effectiveness of Peer Enhanced-computerised Cognitive Behavioural Therapy (P-CCBT) for Adult Subthreshold Depression: Report of the Grant-in-Aid for Scientific Research (no.26350862) by Ministry of Education, Science, Sports and Culture (In Japanese). 2015.
 26. Rush, A.J., 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): p. 573-83.
 27. Fujisawa, D., et al., *Development of Japanese version of QIDS(Quick Inventory of Depressive Symptomatology)-SR(Self-Report)* [in Japanese]. The Japanese Journal of Stress Science, 2010. **25**(1): p. 43-52.
 28. Spitzer, R.L., et al., *A brief measure for assessing generalized anxiety disorder: the GAD-7*. Arch Intern Med, 2006. **166**(10): p. 1092-7.
 29. Muramatsu, K., et al., *Validation and utility of a Japanese version of the GAD-7*. PANMINERVA MEDICA 20th World Congress on Psychosomatic Medicine Abstracts Book, 2009. **51**(Suppl 1 to No 3): p. 79.
 30. Sheehan, K.H. and D.V. Sheehan, *Assessing treatment effects in clinical trials with the discan metric of the Sheehan Disability Scale*. Int Clin Psychopharmacol, 2008. **23**(2): p. 70-83.
 31. Yoshida, T., et al., *Reliability and validity of the Japanese version of the Sheehan Disability Scale* [in Japanese]. Japanese Journal of Clinical Psychopharmacology, 2004. **7**(10): p. 1645-1653.
 32. Attkisson, C.C. and R. Zwick, *The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome*. Evaluation and Program Planning, 1982. **5**(3): p. 233-237.
 33. Prochaska, J.J., et al., *A Therapeutic Relational Agent for Reducing Problematic Substance Use (Woebot): Development and Usability Study*. J Med Internet Res, 2021. **23**(3): p. e24850.
 34. Beatty, C., et al., *Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study*. Front Digit Health, 2022. **4**(847991).
 35. Neimeyer, R.A. and G. Feixas, *The Role of Homework and Skill Acquisition in the Outcome of Group Cognitive Therapy for Depression - Republished Article*. Behav Ther, 2016. **47**(5): p. 747-754.
 36. Mamukashvili-Delau, M., et al., *Long-Term Efficacy of Internet-Based Cognitive Behavioral Therapy Self-Help Programs for Adults With Depression: Systematic Review and Meta-Analysis of Randomized Controlled Trials*. JMIR Ment Health, 2023. **10**: p. e46925.
 37. Fairburn, C.G., et al., *A prospective study of outcome in bulimia nervosa and the long-term effects of three psychological treatments*. Arch Gen Psychiatry, 1995. **52**(4): p. 304-12.
 38. Agras, W.S., et al., *A multicenter comparison of cognitive-behavioral therapy and interpersonal psychotherapy for bulimia nervosa*. Arch Gen Psychiatry, 2000. **57**(5): p. 459-66.
 39. Carter, F.A., et al., *The long-term efficacy of three psychotherapies for anorexia nervosa: a randomized, controlled trial*. Int J Eat Disord, 2011. **44**(7): p. 647-54.
 40. Markowitz, J.C., et al., *Is Exposure Necessary? A Randomized Clinical Trial of Interpersonal Psychotherapy for PTSD*. Am J Psychiatry, 2015. **172**(5): p. 430-40.
 41. Bighelli, I., et al., *Psychosocial and psychological interventions for relapse prevention in schizophrenia: a systematic review and network meta-analysis*. Lancet Psychiatry, 2021. **8**(11): p. 969-980.
 42. Lemmens, L., 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): p. 8-14.
 43. Richards, D. and T. Richardson, *Computer-based psychological treatments for depression: a systematic review and meta-analysis*. Clin Psychol Rev, 2012. **32**(4): p. 329-42.
 44. Rongen, A., et al., *Barriers and facilitators for participation in health promotion programs among employees: a six-month follow-up study*. BMC Public Health, 2014. **14**(1): p. 573.