INTERVENCIÓN Y RESPUESTA

# Digital micro-interventions for depression: what chatbots can and cannot do

SentirIA Research Papers · 2026

Infraestructura de detección temprana en salud mental

Este documento es parte de la base científica de SentirIA,
plataforma de detección temprana y monitoreo continuo de deterioro en salud mental.
No constituye diagnóstico clínico. La evaluación es responsabilidad del profesional.

# Digital micro-interventions for depression: what chatbots can and cannot do

Chatbot-delivered interventions for mild-to-moderate depression produce small-to-moderate effects (Hedges' g = 0.26–0.64 on depression measures), with behavioral activation and cognitive restructuring showing the strongest evidence for brief text-based delivery. These effects are roughly half the magnitude of face-to-face therapy (g = 0.74–0.85) but carry dramatically lower cost and higher scalability. The critical nuance: most techniques produce immediate mood boosts in 2–3 messages, but only sustained engagement over 4–8 weeks generates clinically meaningful symptom reduction. For Latin American populations, where the mental health treatment gap reaches 74.7%, WhatsApp-based and culturally adapted tools like CONEMO and Viva Vida represent the most promising delivery mechanisms, though rigorous chatbot RCTs in the region remain scarce.

---

## Behavioral activation is the strongest candidate for brief digital delivery

Among all therapeutic modalities tested in chatbot format, behavioral activation (BA) has accumulated the most robust evidence. The landmark Woebot RCT (Fitzpatrick, Darcy, & Vierhile, 2017; *JMIR Mental Health*; N=70) demonstrated a between-group effect size of d=0.44 on PHQ-9 scores over just two weeks of daily chatbot interactions versus an information-only control ($F_{1,48}$=6.03, p=.017). More impressively, Mason et al. (2023; *JMIR mHealth and uHealth*; N=103) showed that an 8-week automated SMS program (CBT-txt) achieved d=0.76 for depression, with 53% of participants reaching minimal symptoms versus 15% of waitlist controls. Mediation analysis revealed that 57% of the treatment effect operated through increases in behavioral activation, making BA the single most potent mechanism in text-based depression interventions.

The specific BA components that work in 2–3 text messages include pleasant activity prompts, mood-activity monitoring, and micro-behavioral experiments. Santopetro et al. (2024; *JMIR Mental Health*; N=126) demonstrated that simple daily SMS prompts to complete two enjoyable activities significantly reduced depression versus a passive control over one month—the first fully automated, text-only BA intervention RCT. The StayWell at Home program (2024; *PLOS Digital Health*) further confirmed that single behavioral activation text messages providing "tangible tips for identifying active and pleasurable activities" improved mood in a micro-randomized trial design.

However, several BA components resist brief delivery. Functional analysis (understanding avoidance patterns), values-based activation (linking activities to personal values), and avoidance hierarchies (graded task assignment) require iterative multi-session engagement. Cuijpers et al. (2023; *Psychotherapy Research*; k=22 RCTs, N=819) distinguished four BA types

and found that "pleasant activity" BA (monitoring and scheduling enjoyable activities) is most amenable to digital delivery, while "contextual BA" involving functional analysis and rumination strategies requires structured multi-session protocols.

Face-to-face BA produces larger effects: Hedges' g=0.85 (95% CI: 0.57–1.10) in individual format versus controls (Cuijpers et al., 2023). The COBRA trial (Richards et al., 2016; *The Lancet*; N=440) confirmed BA's non-inferiority to full CBT, with identical 12-month PHQ-9 outcomes (8.4 vs. 8.4) at 20% lower cost. Internet-based BA produces "moderate to large" effects comparable to face-to-face delivery (Alber et al., 2023; *JMIR*; k=13 studies, N=3,274), though Jia et al. (2025; *JMIR*; k=17 RCTs) found digital BA effects were not sustained at 12 months (p=.82).

Null findings matter here. The Tess chatbot tested in Argentina (Bunge et al., 2021; *JMIR Formative Research*; N=181) showed no significant between-group differences at 8 weeks, with ~60% attrition. A Polish adaptation of Woebot (Fido chatbot; Karkosz et al., 2024; *JMIR Formative Research*) also produced null results against a CBT handbook active control—both groups improved equally.

---

## Cognitive restructuring works in single exchanges, but lasting effects require repetition

Laboratory research provides strong evidence that even a single brief cognitive reappraisal exercise produces measurable emotional change. Webb, Miles, and Sheeran's (2012; *Psychological Bulletin*) meta-analysis of 306 experimental comparisons found cognitive reappraisal yielded d+=0.36 for emotional outcomes, with perspective-taking reaching d+=0.45. Neuroimaging meta-analyses (Buhle et al., 2014; *Cerebral Cortex*; 48 studies) confirm that single reappraisal trials reliably modulate bilateral amygdala activity within milliseconds. Critically, depressed individuals show comparable reappraisal benefit to healthy controls when explicitly instructed to reappraise, suggesting the capacity is preserved even when habitual use is diminished.

The most striking evidence for text-based cognitive restructuring comes from Sharma et al. (2024; *CHI 2024*; N=15,531), who deployed an LLM-assisted 5-step CR system on the Mental Health America website. This single cognitive restructuring exchange positively impacted emotional intensity for 67% of participants and helped 65% overcome negative thoughts. The system guided users through describing the thought, detailing the situation, reflecting on emotion, identifying thinking traps with AI assistance, and writing reframes. While this lacks standardized depression outcome measures, the scale is unprecedented.

Within therapy sessions, CR use is a powerful predictor of outcomes. Ezawa and Hollon (2023; *Psychotherapy*; meta-analysis of 4 studies, N=353) found the CR-outcome association was r=.35 (95% CI: .24–.44), equivalent to d=0.85—meaning each individual CR exchange contributes meaningfully to clinical improvement. However, Steinberg et al. (2024; *Journal of Clinical Child &*

*Adolescent Psychology*; N=597 adolescents) tested a single-session digital CR intervention ("Project Think") and found no significant condition differences in symptom reduction versus active control, though CR skill use was maintained at 7-month follow-up. The implication: a single CR session teaches a lasting cognitive skill but may be insufficient alone for symptom change.

Brief cognitive defusion techniques (an acceptance-based alternative to CR) show immediate effects. Barrera et al. (2019; N=68) demonstrated that a 5-minute cognitive defusion exercise immediately reduced thought believability and discomfort with moderate-to-large effect sizes. A VR-delivered single-session cognitive defusion exercise (2025; *JMIR Mental Health*; N=20) produced d=1.01 on defusion measures. Zhu et al. (2025; N=176) showed daily cognitive reappraisal micro-activities over 3 workweeks significantly mitigated negative affect, with effects persisting at 1-month follow-up.

The minimum effective dose for CR remains poorly defined. A scoping review (Pinho et al., 2024) found only 7 studies across all time periods examining how CR should be structured for depression, with only 2 RCTs—a remarkably thin evidence base for one of CBT's core components. Most chatbot meta-analyses find benefits peak after 8 weeks of treatment (He et al., 2023), with effects dissipating at 3-month follow-up (Zhong et al., 2024).

---

## Gratitude interventions show disappointing depression-specific effects

The evidence for gratitude as a depression intervention is considerably weaker than popular enthusiasm suggests. Cregg and Cheavens (2021; *Journal of Happiness Studies*; k=27 studies, N=3,675)—the only meta-analysis focused specifically on depression and anxiety—found gratitude interventions produced just Hedges' g=−0.29 (SE=0.06, p<.01) at post-test and g=−0.23 at follow-up. The authors' own conclusion: "We recommend individuals seeking to reduce symptoms of depression and anxiety engage in interventions with stronger evidence of efficacy."

More damaging, Davis et al.'s (2016; *Journal of Counseling Psychology*; k=26 studies) foundational meta-analysis found that gratitude interventions showed essentially zero advantage over psychologically active comparisons (d=−0.03), collapsing to d=0.02 after trim-and-fill correction for publication bias. Dickens (2017; *Basic and Applied Social Psychology*; k=38 studies) reported the depression-specific effect against neutral controls at just d=0.13 (k=9 studies). Multiple replications of Seligman et al.'s (2005) "Three Good Things" exercise have failed to find depression-specific effects: Mongrain and Anselmo-Matthews (2012; *Journal of Clinical Psychology*; N=411) found PPEs did not exceed control for depression; Gander et al. (2013) found no effect; Sekizawa and Yoshitake (2015) found no difference on CES-D.

Digital gratitude interventions show one critical moderator: they work only for already-distressed individuals. A 2025 app study (*JMIR mHealth and uHealth*; N=120) found d=−0.68 (p=.04) for

participants with moderate-or-higher baseline symptoms, but d=0.16 (p=.46) for the full sample—a non-significant effect. Heckendorf et al. (2019; *Behaviour Research and Therapy*; N=260) found an internet/app-based gratitude intervention reduced depression through reducing repetitive negative thinking (d=0.61 at post-test, d=0.75 at 3-month follow-up), suggesting the mechanism may be cognitive rather than affective.

For duration, the evidence suggests 4+ weeks of repeated practice is necessary for meaningful depression effects. Bohlmeijer et al. (2021; *Frontiers in Psychology*; N=169) found gratitude mood mediated well-being effects at 4 weeks but not at 2 weeks. Lyubomirsky et al. (2005) found counting blessings once per week was more effective than three times per week, which actually decreased well-being—likely due to adaptation. No published RCT has tested a chatbot specifically delivering gratitude exercises for depression as a primary outcome, representing a clear research gap.

---

## Mindfulness micro-doses may work, but the evidence is thinner than expected

App-based mindfulness interventions produce consistent small effects on depression. Gál, Ștefan, and Cristea (2021; *Journal of Affective Disorders*; 34 RCTs, N=7,566) found g=0.33 (95% CI: 0.24–0.43) for depression. The updated Linardon et al. (2024; *Clinical Psychology Review*; 45 RCTs) found g=0.24 (95% CI: 0.17–0.31, NNT=13.57) for depression, though mindfulness apps showed no advantage over active therapeutic comparisons (g=−0.15, ns). Full 8-week MBSR/MBCT programs produce considerably larger effects: d=0.55 versus no treatment and d=0.71 versus non-specific controls (Goldberg et al., 2018, 2019; *Clinical Psychology Review*; 142 RCTs, N=12,005).

The dose-response relationship is surprisingly flat. Strohmaier (2020; *Mindfulness*; 203 RCTs, N=15,971) found no statistically significant dose-response effects for depression, anxiety, or stress in meta-regression. Most provocatively, Strohmaier, Jones, and Cane (2021; *Mindfulness*; N=71) found that 5-minute sessions produced significantly greater improvements in trait mindfulness (d=2.17, p<.001) and stress (d=−1.18, p<.01) than 20-minute sessions for novice meditators. Schillings et al. (2023; *Scientific Reports*; N=372) confirmed 10 minutes was as effective as 20 minutes. These findings challenge "more is better" assumptions, though they measured state mindfulness rather than depression specifically.

For ultra-brief mindfulness (1–3 minutes), direct evidence on depression is sparse. Broderick (2005) showed an 8-minute guided exercise reduced negative affectivity after mood induction. Howarth et al.'s (2019; *Mindfulness*) systematic review of 85 studies found that brief MBIs of 5–20 minutes in a single session positively impacted negative mood and anxiety. But Schumer, Lindsay, and Creswell's (2018; *Journal of Consulting and Clinical Psychology*; 65 RCTs, N=5,489) meta-analysis of brief mindfulness training found a small overall effect of g=0.13 (p=.001) on

negative affectivity that collapsed to g=0.04 after publication bias correction—a sobering finding.

Text-message-delivered mindfulness specifically has weak evidence. Senanayake et al. (2019; *Journal of Telemedicine and Telecare*; 7 RCTs, N=845) found a marginally non-significant pooled effect of SMD=0.23 (95% CI: −0.02 to 0.48) for text messaging interventions on depression. The single strongest result comes from Agyapong et al. (2017; *BMC Psychiatry*; N=73), where twice-daily supportive text messages for 3 months produced d=0.67, though this combined mindfulness with psychoeducation and CBT content.

---

## Single-session interventions produce real but small effects that often fade

Jessica Schleider's research program provides the most rigorous evidence on single-session interventions (SSIs). Her foundational meta-analysis (Schleider & Weisz, 2017; *JAACAP*; 50 RCTs, N=10,508 youth) found an overall SSI effect of Hedges' g=0.32, but the depression-specific effect was just g=0.21 and non-significant (only 6 trials). Youth-focused CBT approaches yielded the largest effects at g=0.74. SSIs versus no-treatment controls produced g=0.41, but versus active controls only g=0.14.

Her largest trial—a COVID-era 3-arm RCT (Schleider et al., 2022; *Nature Human Behaviour*; N=2,452 adolescents)—compared BA-SSI and Growth Mindset SSI to supportive control. Both active SSIs reduced 3-month depression at d=0.18, matching the meta-analytic estimate precisely. The growth mindset SSI from her earlier work (Schleider & Weisz, 2018; *JCPP*; N=96) showed effects strengthening over 9 months (d=0.60 for parent-reported depression), an unusual finding suggesting mechanism-targeted SSIs may produce cascading benefits.

The most sobering data comes from the Kaveladze et al. (2026; *Nature Human Behaviour*) megastudy: N=7,505 adults randomized to 12 different digital SSIs (each ≤10 minutes) with 4-week follow-up. Nearly all SSIs improved hope and agency immediately (d≤0.37), but only 2 of 12 SSIs significantly reduced depression at 4 weeks: interactive cognitive reappraisal (Koko platform, d=0.15) and Finding Focus (d=0.14). The validated behavioral activation SSI did not reduce 4-week depression in adults. Some SSIs paradoxically decreased readiness to change. The critical design insight: "The most successful SSIs provided focused, engaging, and actionable guidance on a skill that directly addressed users' struggles."

Comparing SSIs to multi-session digital interventions, the effect size gap is substantial. Multi-session digital interventions produce approximately g=0.52 overall (Moshe et al., 2021; 83 studies, N=15,530), with guided interventions reaching g=0.63 versus g=0.34 for self-help. SSIs produce roughly one-third to one-half this magnitude. Schleider's 2025 umbrella review (*Annual Review of Clinical Psychology*; 24 systematic reviews, 415 trials) found a pooled SMD of −0.25 across 12 meta-analytic reviews.

The IntelliCare suite (Mohr et al., 2017; *JMIR*; N=99) offers an alternative model—13 micro-apps with average interaction time of 1.1 minutes, used 3–4 times daily over 8 weeks. This produced ~50% symptom reduction in a single-arm trial, though the lack of control limits interpretation. The RCT (Mohr et al., 2019; N=301) confirmed significant improvement across all arms. This "micro-dose, high-frequency" approach may outperform single concentrated sessions.

---

## What the chatbot meta-analyses actually show

The chatbot evidence base has matured substantially, with multiple meta-analyses now available. Effect sizes vary considerably by analysis:

- Zhong et al. (2024; *Computers in Human Behavior*; 18 RCTs, N=3,477): g=−0.26 (95% CI: −0.34 to −0.17) for depression—the most conservative and methodologically careful estimate. No heterogeneity detected; no publication bias. Effects strongest at 8 weeks, not sustained at 3-month follow-up.
- Li et al. (2023; *npj Digital Medicine*; 15 RCTs): g=0.64 (95% CI: 0.17–1.12) for depression—moderate-to-large but with wide confidence intervals indicating substantial heterogeneity.
- Linardon et al. (2024; 176 app RCTs): Chatbot-specific subgroup showed g=0.53 (95% CI: 0.33–0.74) for depression, significantly larger than the overall app effect of g=0.28.
- He et al. (2023; *JMIR*; 32 RCTs): Depression effects ranged g=0.24–0.62 depending on outcome specification.
- Peretz et al. (2025; *npj Digital Medicine*; 39 studies, N=7,401): g=0.31 (95% CI: 0.17–0.46) for depression, with larger effects in clinical/subclinical versus nonclinical samples.

Therapeutic alliance with chatbots is surprisingly robust. Darcy et al. (2021; *JMIR Formative Research*; N=36,070 Woebot users) found WAI-SR bond subscale scores of 3.84/5.0 (SD=1.0), comparable to individual face-to-face CBT (mean ~4.0) and group CBT (~3.8). Wysa showed similar bond scores of 3.98/5.0 (Beatty et al., 2022; *Frontiers in Digital Health*; N=1,205). Alliance formed within 3–5 days of first use and remained stable. Chatbot design features predict outcomes: multimodal chatbots (voice + text + animations) significantly outperform text-only (Liu et al., 2025; N=84; adherence d=0.82), and AI-based chatbots (g=−0.36) outperform rule-based systems (g=−0.09).

The first RCT of a fully generative AI chatbot (Therabot; Heinz et al., 2025; *NEJM AI*; N=210) produced substantially larger effects: d=0.845–0.903 for MDD (PHQ-9 change: −6.13 vs. −2.63 at 4 weeks). Therapeutic alliance was rated comparable to human therapists. No serious adverse events occurred. However, this single trial requires replication, and LLM-based chatbots carry unique risks: hallucinations, "toxic positivity," unsolicited advice-giving, and crisis recognition failures.

Engagement patterns are critical: average attrition across chatbot studies is approximately 21% (Jabir et al. meta-analysis), somewhat better than general mental health apps (~25%). Dropout peaks in the first week, and engagement tends to decline after 2 weeks. Push notifications, personalized messages, and deployment on familiar platforms (WhatsApp, Facebook Messenger) significantly improve adherence.

---

## The stepped care model and knowing when to escalate

Chatbots fit naturally as Tier 1 interventions in stepped care frameworks. The UK's NHS Talking Therapies system has integrated Limbic Access as a self-referral/triage chatbot, and NICE is actively evaluating Wysa for similar functions. An umbrella review of stepped care effectiveness (2024; PROSPERO CRD42023461710; 10 systematic reviews, 38 primary studies) found stepped care improved depression response at 3–6 months (RR=1.52, 95% CI: 1.30–1.78) and 9–12 months (RR=1.47, 95% CI: 1.23–1.77).

Evidence-based escalation criteria include PHQ-9 scores ≥20 (severe depression), persistent suicidal ideation, active psychosis, and failure to improve after initial chatbot engagement. Wysa's AI system detects 82% of crisis instances (confirmed by users), triggering SOS interventions with safety planning and crisis helplines (data from N=19,950 users; 5.2% reported crisis instances). NLP-based suicide risk detection achieves 72–93% accuracy across systematic reviews using social media and health record data (Bejan et al., 2022). Transformer-based models (BERT, RoBERTa) outperform traditional ML approaches.

Blended care consistently outperforms standalone chatbot use. Enhanced e-CBT shows approximately 30% lower attrition than unguided digital interventions, and meta-analyses consistently find larger effects when digital components include therapist support (guided g≈0.63 versus self-help g≈0.34). Wysa reportedly handles 80% of the support load, allowing human therapists to concentrate on complex cases—a task-shifting model particularly relevant for resource-limited settings.

---

## Which acute-effect techniques have the strongest digital evidence

Among specific techniques with RCT evidence for digital delivery, the evidence hierarchy for acute effects is:

- Cognitive reappraisal has the fastest documented onset (measurable neural effects within milliseconds; behavioral effects with d+=0.36 in single lab exercises; Webb et al., 2012) and was one of only two SSIs to maintain depression effects at 4 weeks in the Kaveladze

megastudy (d=0.15).

- Self-compassion exercises show moderate-to-large effects in digital packages. Levin et al. (2014; N=76) found web-based ACT with values and self-compassion components produced d=0.97 for depression. Mobile self-compassion apps show significant effects (29k app RCT, 2024; N=63; significant depression reduction versus waitlist).
- Values clarification produces d=0.41 alone and d=1.37 combined with mindfulness in lab-based exercises (Levin, Hildebrandt, Lillis, & Hayes, 2012). Web-based ACT programs integrating values show large effects (d=0.69–0.97 for depression).
- Cyclic sighing/breathing exercises show the fastest physiological onset—5 minutes to measurable mood improvement (Balban et al., 2023, Stanford).
- Sleep hygiene alone is consistently outperformed by comprehensive digital CBT-I in RCTs. Digital CBT-I produces SMD=−0.42 (95% CI: −0.56 to −0.28) for depression (meta-analysis of 22 RCTs, *npj Digital Medicine*, 2023), but sleep hygiene education serves as the weaker active control.
- Social connection prompts lack isolated digital RCTs. Hansen et al. (2025; 40 RCTs, N=6,062) found self-guided social interventions had limited impact on loneliness.
- Psychoeducation alone produces modest effects (typically used as active control) but is a useful scaffolding component.

Elefant et al. (2017; *Internet Interventions*; N=122) tested three CBT micro-interventions (breathing exercise, thought record, pleasant activities selector) and found significant immediate mood improvements but no lasting effects at 4-week follow-up. This encapsulates the core challenge: micro-interventions improve mood acutely but may lack "therapeutic potency" for lasting change as standalone tools.

---

## Latin America's treatment gap demands culturally adapted digital solutions

Latin America faces a mental health treatment gap of 74.7% (Kohn et al., 2018; *Revista Panamericana de Salud Pública*), with a mental health workforce of just 8.7 per 100,000 compared to 125.2 in the United States. Less than 2% of health budgets are allocated to mental health, and up to 90% of that goes to psychiatric hospitals rather than community care. Over 16 million adolescents aged 10–19 in Latin America live with a mental health condition.

The largest and most rigorous LATAM evidence comes from the CONEMO trials (Araya et al., 2021; *JAMA*; N=880 in São Paulo, N=432 in Lima). This smartphone-delivered BA intervention—18 sessions over 6 weeks, each under 10 minutes, minimally supported by nurse assistants—achieved OR=1.6 (95% CI: 1.2–2.2) for 50% PHQ-9 reduction in Brazil and OR=2.1 (95% CI: 1.4–3.2) in Peru. However, effects were not sustained at 6 months in Brazil. The Viva Vida WhatsApp RCT (Scazufca et al., 2024; *Nature Medicine*; N=603 older adults in São Paulo)

showed depression improved in 42.4% of intervention versus 32.2% of control using self-help audio and visual WhatsApp messages—notable because most participants had never received any mental health treatment.

Cultural adaptation frameworks are essential. Bernal, Bonilla, and Bellido's (1995) Ecological Validity Framework identifies eight dimensions (language, metaphors, goals, concepts, persons, content, context, methods), and this remains the gold standard. Key cultural values to integrate include familismo (deep family commitment; meta-analytically protective against depression per Valdivieso-Mora et al., 2016), personalismo (warm personal relationships), and respeto. BA may be more culturally consonant than traditional CBT for Latino populations because it attributes depression to environmental reward depletion rather than internal cognitive deficits, which can trigger self-stigma.

The only purpose-built Spanish-language BA app tested in an RCT is ¡Aptívate! (Dahne et al., 2019; *Journal of Affective Disorders*; N=42 Latinx adults with limited English proficiency), which incorporated values-based activity scheduling with familismo-consistent activities. Wysa's Spanish adaptation (Dinesh et al., 2024; *Digital Health*; N=2,767 Spanish-speaking users) showed significantly longer conversations (d=0.44), more sessions (d=0.18), and greater distress disclosure in Spanish versus English—highlighting that digital tools must be adapted beyond simple translation. StayWell at Home found Latinx users showed an additional −1.45 point PHQ-8 decline compared to non-Latinx White users (2023; *Behaviour Research and Therapy*; N=398).

---

## Conclusion: a pragmatic hierarchy for chatbot-delivered depression interventions

The evidence points to a clear hierarchy for chat-based micro-interventions. Behavioral activation—specifically pleasant activity scheduling and mood-activity monitoring—is the single best-supported technique for brief text delivery, with effect sizes of d=0.44–0.76 in RCTs and confirmed mediation through activation mechanisms. Cognitive reappraisal ranks second, with strong laboratory evidence for single-exchange effects (d=0.36) and emerging real-world data from LLM-assisted platforms. Mindfulness and gratitude produce smaller and less reliable depression-specific effects, though they serve useful adjunctive roles.

Three design principles emerge from the evidence. First, frequency beats intensity: the IntelliCare model (1-minute interactions, 3–4 times daily) and daily SMS prompts outperform single concentrated sessions. Second, sustained engagement is non-negotiable—virtually all chatbot interventions require 4–8 weeks to produce meaningful depression symptom change, and effects consistently fade at 3-month follow-up. Third, multimodal and AI-driven chatbots significantly outperform text-only rule-based systems, with the Therabot generative AI trial (d=0.85–0.90) suggesting a possible step-change, though replication is urgently needed.

The "without replacing a therapist" framing is well-supported by stepped care evidence. Chatbots

work best as Tier 1 interventions for mild-to-moderate symptoms, with built-in escalation protocols for severe symptoms (PHQ-9≥20), suicidality, and treatment non-response. The most impactful deployment model combines chatbot self-management with periodic therapist oversight—a blended approach that is especially critical for Latin American contexts where the 74.7% treatment gap and 8.7-per-100,000 mental health workforce make pure therapist-delivered care structurally impossible at scale.