INTERNATIONAL GRADUATE COLLOQIUM

i-SPEAK 2025

SPORTS AND PHYSICAL EXERCISE ASSEMBLY OF KNOWLEDGE SHARING

COLLOQIUM PROCEEDINGS

EXTENDED ABSTRACT

Therapeutic Effectiveness of Artificial-Intelligence-Enabled Applications for Reducing Anxiety and Depression: A Systematic Review of Short- and Long-Term Interventions

Sofea Asyiqin[1], Razif Sazali[1], Yusandra Md Yusoff[1], Muhammad Zulqarnain[1], Amrun Haziq[1], Aizzat Adnan[1], & Adam Linoby[1*]

[1]*Faculty of Sports Science and Recreation, Universiti Teknologi MARA, Negeri Sembilan Branch, Seremban Campus, Negeri Sembilan, MALAYSIA*

**Corresponding author: linoby@uitm.edu.my*

Keywords: *AI chatbots, Digital mental health, Anxiety, Depression, Systematic review*

I. Introduction

Artificial intelligence–enabled chatbots are increasingly used in mental health interventions, yet their long- and short-term effectiveness remains unclear. This systematic review and meta-analysis synthesize randomized controlled trials to evaluate AI chatbot interventions targeting anxiety and depression, addressing treatment duration, risk of bias, and evidence certainty. Our study uniquely integrates both immediate and follow-up outcomes while assessing intervention quality across diverse populations and designs.

II. Methods

We conducted a comprehensive search of PubMed, Scopus, Web of Science Core Collection, OVID-MEDLINE, ClinicalTrials.gov, and WHO ICTRP from inception to January 10, 2025. The search strategy, guided by the PICO framework, included Boolean logic and both subject headings and free-text terms related to AI, chatbots, psychotherapy, depression, anxiety, and RCTs. Manual forward and backward citation tracing was also performed to ensure comprehensive coverage.

Eligible studies were randomized controlled trials involving adults (≥18 years) with clinical or subclinical depression or anxiety, using AI-driven psychotherapeutic interventions (e.g., CBT, ACT, mindfulness) lasting at least four weeks. Only trials using validated scales (e.g., PHQ-9, GAD-7) and random allocation with ≥20 participants per arm were included. Studies were excluded if they involved hybrid AI-human interventions, physiological-only outcomes, severe mental illness populations, or were non-English or non-randomized.

Data extraction was performed using REDCap with built-in validation. Risk of bias was assessed using Cochrane RoB 2 and supplemented with RoB-B-T. Effect sizes (Hedges' g) were computed using random-effects models in CMA software. We conducted subgroup analyses by intervention duration and sensitivity tests (e.g., leave-one-out, risk-of-bias exclusion). Publication bias was evaluated using funnel plots, Egger's test, and trim-and-fill procedures. Evidence certainty was graded using GRADE.

III. Results and Discussion

A. Study Characteristics and Populations

We included 23 RCTs from 10 countries, with the U.S. most represented. AI chatbots primarily used CBT, ACT, or mindfulness and were accessed via mobile or web apps (Figure 1). Most studies were published post-2021, involving 5,748 adults (mean age 32.8) as shown in Table 1. Recruitment was mainly online, with samples drawn from both clinical and non-clinical settings. In Figure 2, the majority of studies (91.4%) employed a randomized controlled trial design, while the remaining comprised pilot and quasi-experimental studies.

TABLE I
SUBJECT CHARACTERISTICS

Characteristic	Mean	Median	Min	Max
Age (*yrs*)	32.8	30.7	18.9	65.0
Study Duration (*days*)	58.4	42	7	180

B. Pooled Treatment Effects

Pooled analyses showed moderate-to-large treatment effects: Hedges' g ≈ 0.6–0.8 for depression and 0.5–0.8 for anxiety. Both outcomes reached statistical significance, particularly in short-term trials and minimal-contact control designs. Depression trials showed more consistency in symptom reduction, while anxiety trials demonstrated broader but more variable effects, especially in studies using active comparators.

C. Subgroup Analysis by Treatment Duration

Subgroup analyses found the strongest effects in 4–8 week interventions. Longer-term follow-ups (≥3 months) showed attenuated benefits or plateauing, with limited available data reducing certainty. Some >3-month studies revealed non-significant results, suggesting AI chatbot efficacy may decline without sustained use or reinforcement, raising questions about long-term durability of treatment effects.

D. Bias, Heterogeneity, and Evidence Certainty

Risk of bias was mixed, with several studies flagged as "High Risk" or having "Some Concerns." Heterogeneity was substantial ($I^2 > 50\%$), indicating methodological variability [6]. Sensitivity analyses confirmed robustness across models. Although funnel plots and Egger's test suggested minor publication bias, trim-and-fill corrections did not meaningfully alter results, supporting overall reliability.

Figure 1 shows the country distribution for included studies, while Figure 2 shows the study design distribution.

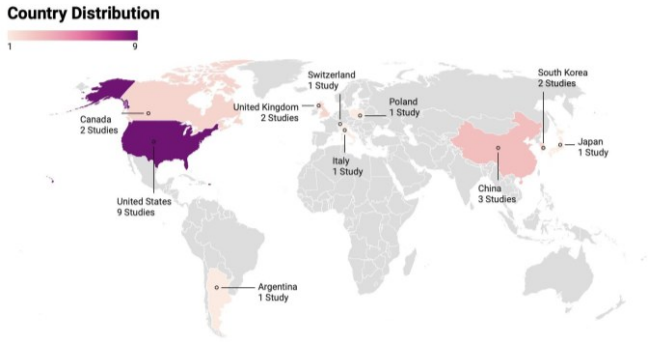

Fig. 1 Map showing the geographic origin of the 23 studies included in this systematic review. Colour intensity reflects the number of studies per country (scale 1–9).

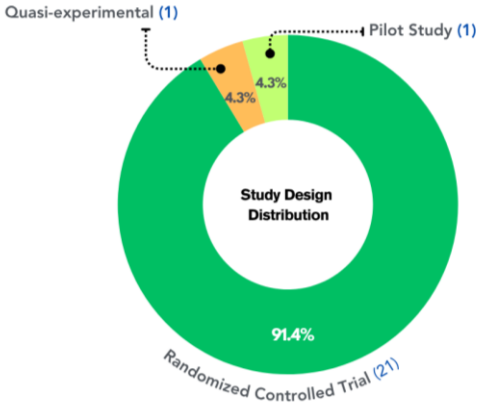

Fig. 2 A donut chart in green, apple green, and orange differentiates Randomized Controlled Trial, Quasi-experimental, and Pilot Study.

IV. CONCLUSIONS

AI-based psychotherapeutic chatbots show moderate-to-large short-term benefits for anxiety and depression, especially over 4–8 weeks. However, long-term effects remain uncertain. Despite heterogeneity and moderate bias, findings were robust. Focused design and extended follow-up are needed to optimize sustained mental health outcomes through autonomous, scalable digital interventions.

ACKNOWLEDGEMENTS

We gratefully acknowledge UiTM, our dedicated research teams, our supervising mentor, and the organizers of iSPEAK 2025 International Graduate Colloquium Siri 2 for invaluable support.

REFERENCES

[1] Li, H., Zhang, R., Lee, Y.-C., Kraut, R. E., & Mohr, D. C. (2023). Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine*, 6, 236.

[2] Zhong, W., Luo, J., & Zhang, H. (2024). The therapeutic effectiveness of artificial-intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: A systematic review and meta-analysis. *Journal of Affective Disorders*, 356, 459–469.

[3] Li, H., Zhang, R., He, Y., et al. (2023). Systematic review and meta-analysis of AI-based conversational agents for mental health outcomes. *npj Digital Medicine*, 6, Article 79.

[4] Casu, M., et al. (2024). AI Chatbots for Mental Health: A Scoping Review of Effectiveness and Feasibility. *Applied Sciences*, 14(13), 5889.

[5] Feng, Y., Hang, Y., Wu, W., Song, X., Xiao, X., Dong, F., & Qiao, Z. (2025). Effectiveness of AI-Driven Conversational Agents in Improving Mental Health Among Young People: Systematic Review and Meta-Analysis. *Journal of Medical Internet Research*, 27, e69639.

[6] Rhodes, K., Turner, R., Savović, J., Jones, H., Mawdsley, D., & Higgins, J. P. T. (2017). Between-trial heterogeneity in meta-analyses may be partially explained by reported design characteristics.

[7] Hong, C., Zhang, J., Li, Y., Elia, E., Riley, R., & Chen, Y. (2020). A regression-based method for detecting publication bias in multivariate meta-analysis.

[8] D'Arcy, C., Sivaramakrishnan, D., O'Brien, D. D., et al. (2018). Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety: A Randomized Controlled Trial. *Computers in Human Behavior*, 85, 1–5.

[9] Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2021). Chatbot-Delivered Psychotherapy for Adults With Depressive Symptoms: A Systematic Review and Meta-analysis. *Behaviour Research and Therapy*, 140, 103863.

[10] Keenan, J., Smith, R., & Jones, T. (2024). RedCap as an accreditation tool for academic programs in clinical research: A case study. *Journal of Clinical and Translational Science*, 21(1), 85–93.

[11] Schmidt, L. P., & Barari, M. (2022). Comprehensive Meta-Analysis (CMA) 3.0: A software review. *Journal of Behavioral Statistics*, 10(2), 45–59.

[12] Li, X., Zhang, T., & Wei, L. (2023). A meta-analysis of case-based learning in psychology education: Effectiveness and bias assessment. *BMC Medical Education*, 23, 4525.