Research Protocol: Testing an AI emotional support companion for university students' psychological wellbeing

Project Summary:
University students often experience mild-to-moderate psychological distress, yet many do not access traditional mental health support due to stigma, limited resources, or accessibility barriers. This randomized controlled trial assesses the effectiveness of Kai, an AI-based mental health and emotional support companion, compared to face-to-face therapy and a waitlist control. Two experiments will be conducted: Experiment 1 compares Kai to group therapy and a waitlist; Experiment 2 compares Kai to individual counseling and a waitlist. Each intervention runs for 12 weeks. Mental health outcomes—life satisfaction, anxiety, depression, PTSD symptoms, well-being, and substance use—will be measured pre- and post-intervention. The study will also evaluate mediating (engagement, therapeutic alliance) and moderating factors (attachment style, social support, loneliness, political stress). Conducted at Reichman University, this trial will provide insights into the potential of AI to deliver scalable mental health support.

General Information
Protocol Title: Effectiveness of Kai AI Companion for Student Wellbeing: A Randomized Controlled Trial
Protocol Number: p_2024178
Date: 08/05/2025

Sponsor:
Reichman University | Research Authority | Public Benefit Company P.O.Box 167, Herzliya 4610101, Israel | www.runi.ac.il | Tel: + 972 9 9502426 | Fax: +972 9 9527268 research@runi.ac.il

Contact:
Principal Investigator:
Prof. Anat Shoshani
Baruch Ivcher School of Psychology, Reichman University
Prof. Anat Shoshani | +972 9 9602887 | ashoshani@runi.ac.il

Rationale & Background Information

AI-driven mental health tools are revolutionizing digital support by offering scalable, always-on, and personalized interventions. Kai, an AI emotional support companion, represents this evolution. Designed to simulate supportive conversations, Kai uses natural language processing and evidence-based frameworks (CBT, ACT, and positive psychology) to deliver empathetic, responsive, and contextually aware support. Research has shown that users often engage meaningfully with AI companions, reporting therapeutic benefits akin to human interaction (Boucher et al., 2021; Vaidyam et al., 2019).

The CASA paradigm (Nass & Moon, 2000) suggests that individuals attribute social characteristics to AI, fostering trust and openness. AI's perceived empathy, responsiveness, and humanness are associated with increased user engagement and improved outcomes (Chaudhry & Debi, 2024; Limpanopparat et al., 2024). Systematic reviews support the clinical efficacy of AI conversational agents for depression and anxiety symptom reduction (Li et al., 2023). AI companions are also valuable in contexts of high psychological vulnerability, such as conflict or pandemic-related stress, where timely human support is unavailable (Charlson et al., 2019; Carpiniello, 2023).

Moderators such as attachment style (Hsieh & Oh, 2024), social support (Zimet et al., 1990), and loneliness (Gabarrell-Pascuet et al., 2023) influence how individuals interact with digital agents. Loneliness and perceived isolation may both increase engagement and modulate outcome responsiveness. . Furthermore, AI companionship may alter help-seeking behaviors, either complementing or substituting traditional therapy (Xie & Pentina, 2022). This study seeks to understand the effectiveness, moderators, and mechanisms of change within AI-delivered psychological support and to inform scalable mental health strategies that can extend care to underserved and high-risk groups.

Study Goals and Objectives
Goals:
Evaluate Kai's effectiveness in improving student mental health.

Primary Objectives:
1. Assess improvement in mental health outcomes vs. waitlist.
2. Compare Kai with group and individual therapy.

Secondary Objectives:
- Evaluate effects on therapy-seeking intention.
- Evaluate effects on substance use,
- Examine moderators (attachment, support, stress).
- Investigate mediators (engagement, perceived AI qualities).

Study Design Type: Two three-arm parallel RCTs
Design: Interventional, randomized, single-blind (assessor), parallel assignment
Setting: University-based
Sample Size: Exp 1: 984; Exp 2: 164 (total: 1148)
Statistical Power: 90%, $\alpha = 0.05$ (medium effect size, allowing for 20% attrition)

Instruments
Measures will be collected at baseline and again after 3 months.
- GAD-7 (Spitzer et al., 2006)
- PHQ-9 (Kroenke et al., 2001)
- WHO-5 (Sischka et al., 2020)
- PCL-5 (Price et al., 2016)
- ASSIST (WHO)
- BMSLSS (Seligson et al., 2003)
- General Help-Seeking Questionnaire
- Perceived Social Support (Zimet et al., 1990)
- Attachment Style (Lafontaine et al., 2016)
- Loneliness Scale (Hughes et al., 2004)

Methodology
Experiment 1:
- Group 1: Kai AI Companion (use recommended ≥3x/week, 10–15 min/session, but flexible according to participant preference; 12 weeks)
- Group 2: Group Therapy (12 weekly 90-minute sessions with a clinical psychologist)
- Group 3: Waitlist Control (offered access after study)

Experiment 2:
- Group 1: Kai AI Companion (same as above)
- Group 2: University Counseling Services (individual sessions, weekly for 3 months)
- Group 3: Waitlist Control

Recruitment will be conducted via digital flyers, email lists, and student networks. Randomization will be performed using a computer algorithm. Allocation will remain concealed. Assessors will be single-blinded. Data will be collected digitally.

Safety Considerations

Participants will be screened to exclude acute risk. The safety protocol will include weekly distress monitoring, emergency contact lists, and real-time escalation for suicidal ideation. All staff will be trained in referral procedures. Consent forms will highlight potential emotional triggers. Adverse events will be recorded and reviewed bi-weekly. A licensed clinical psychologist will be on call for participants in crisis.

Data Management & Statistical Analysis

Data will be anonymized and stored on encrypted servers. It will be coded using unique IDs. SPSS and AMOS will be used for data analysis. Repeated measures MANOVA with covariates will be conducted. Moderation will be assessed via interaction terms (PROCESS). Mediation models will explore indirect effects via engagement and therapeutic alliance. Missing data will be handled using multiple imputation. All variables will be checked for assumptions, and sensitivity analyses will be conducted.

Quality Assurance

Standardized facilitator manuals, digital checklists, and fidelity ratings will be implemented. Monthly audit meetings will be held. Ethics will be monitored by the university IRB. Clinical monitors will ensure adherence to GCP.

Expected Outcomes

Study Hypotheses

1. Primary Hypotheses:
    a. University students using the Kai AI companion will show significantly greater improvements in anxiety, depression, and wellbeing compared to waitlist controls.
    b. The effectiveness of the Kai AI companion for improving psychological outcomes will be comparable to or exceed face-to-face group and interpersonal therapy across multiple mental health domains.
    c. Kai AI intervention will significantly reduce substance use behaviors compared to both face-to-face group therapy and waitlist control conditions.
    d. Users of the Kai AI companion will report decreased intention to seek traditional therapy compared to participants in the face-to-face therapy and waitlist control groups.
2. Moderation Hypotheses:
    a. The effectiveness of the Kai AI intervention will be moderated by relational factors, with participants reporting higher loneliness, lower social support, and higher attachment insecurity (both anxious and avoidant) showing stronger intervention benefits, particularly for anxiety outcomes.
    b. Political life events exposure will moderate intervention effectiveness, with participants reporting higher exposure to political stressors showing more substantial benefits from Kai AI intervention, especially for anxiety and depression outcomes.
3. Mediation and Engagement Hypotheses:
    a. Higher engagement with Kai (measured by message frequency) will be associated with greater improvements in mental health outcomes.
    b. Therapeutic alliance (perceptions of Kai's professional competence and empathy) will predict user engagement, which will in turn mediate mental health improvements.
    c. Mental health improvements through Kai usage will mediate reductions in perceived need for traditional therapy.

We expect that Kai will produce equal or superior results to traditional therapy in improving psychological wellbeing. The study will identify profiles of students who benefit most from AI support. Outcomes will inform ethical AI use in mental health and the development of hybrid care models combining digital and human support.

References

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., Parks, A., & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: A review. Expert Review of Medical Devices, 18(sup1), 37–49. https://doi.org/10.1080/17434440.2021.2013200

Carpiniello, B. (2023). The mental health costs of armed conflicts—a review of systematic reviews conducted on refugees, asylum-seekers and people living in war zones. International Journal of Environmental Research and Public Health, 20(4), 2840. https://doi.org/10.3390/ijerph20042840

Charlson, F., van Ommeren, M., Flaxman, A., Cornett, J., Whiteford, H., & Saxena, S. (2019). New WHO prevalence estimates of mental disorders in conflict settings: A systematic review and meta-analysis. The Lancet, 394(10194), 240-248. https://doi.org/10.1016/s0140-6736(19)30934-1

Chaudhry, B. M., & Debi, H. R. (2024). User perceptions and experiences of an AI-driven conversational agent for mental health support. Mhealth, 10, 22. https://doi.org/10.21037/mhealth-23-55

Gabarrell-Pascuet, A., García-Mieres, H., Giné-Vázquez, I., Moneta, M. V., Koyanagi, A., Haro, J. M., & Domènech-Abella, J. (2023). The association of social support and loneliness with symptoms of depression, anxiety, and posttraumatic stress during the COVID-19 pandemic: A meta-analysis. International Journal of Environmental Research and Public Health, 20(4), 2765. https://doi.org/10.3390/ijerph20042765

Hsieh, I. T., & Oh, C. H. (2024). Users' attachment styles and ChatGPT interaction: Revealing insights into user experiences. Journal of the Korea Society of Computer and Information, 29(3), 21-41.

Hughes, M. E., Waite, L. J., Hawkley, L. C., & Cacioppo, J. T. (2004). A short scale for measuring loneliness in large surveys: Results from two population-based studies. Research on Aging, 26, 655-672. https://doi.org/10.1177/0164027504268574

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. Journal of General Internal Medicine, 16(9), 606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lafontaine, M.-F., Brassard, A., Lussier, Y., Valois, P., Shaver, P. R., & Johnson, S. M. (2016). Selecting the best items for a short-form of the Experiences in Close Relationships questionnaire. European Journal of Psychological Assessment, 32, 140-154. https://doi.org/10.1027/1015-5759/a000243

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(1), 236. https://doi.org/10.1038/s41746-023-00979-5

Limpanopparat, S., Gibson, E., & Harris, A. (2024). User engagement, attitudes, and the effectiveness of chatbots as a mental health intervention: A systematic review. Computers in Human Behavior: Artificial Humans, 100081. https://doi.org/10.1016/j.chbah.2024.100081

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81-103. https://doi.org/10.1111/0022-4537.00153

Price, M., Szafranski, D. D., van Stolk-Cooke, K., & Gros, D. F. (2016). Investigation of abbreviated 4 and 8 item versions of the PTSD Checklist 5. Psychiatry Research, 239, 124-130. https://doi.org/10.1016/j.psychres.2016.03.014

Seligson, J. L., Huebner, E. S., & Valois, R. F. (2003). Preliminary validation of the brief multidimensional students' life satisfaction scale (BMSLSS). Social Indicators Research, 61, 121-145. https://doi.org/10.1023/A:1021326822957

Sischka, P. E., Costa, A. P., Steffgen, G., & Schmidt, A. F. (2020). The WHO-5 well-being index–validation based on item response theory and the analysis of measurement invariance across 35 countries. Journal of Affective Disorders Reports, 1, 100020. https://doi.org/10.1016/j.jadr.2020.100020

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: a review of the psychiatric landscape. The Canadian Journal of Psychiatry, 64(7), 456-464. https://doi.org/10.1177/0706743719828977

Xie, T., & Pentina, I. (2022). Attachment theory as a framework to understand relationships with social chatbots: A case study of Replika.

Zimet, G. D., Powell, S. S., Farley, G. K., Werkman, S., & Berkoff, K. A. (1990). Psychometric characteristics of the multidimensional scale of perceived social support. Journal of Personality Assessment, 55(3-4), 610-617. https://doi.org/10.1080/00223891.1990.9674095