

Breaking Barriers in Student Mental Health Care With AI-Enhanced Group Cognitive Behavioral Therapy: A Pilot Feasibility Study

Bogdan Tudor Tulbure, Bianca Spătaru, Andrei Rusu, Darian Faur, Daniel Precupaș, Loredana Marcela Trancă, Șerban Gabriel Popa, Adrian Spătaru

Submitted to: JMIR Formative Research
on: September 17, 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

Supplementary Files..... 40

 Figures 41

 Figure 1..... 42

CONSORT (or other) checklists..... 43

 CONSORT (or other) checklist 0..... 44

Breaking Barriers in Student Mental Health Care With AI-Enhanced Group Cognitive Behavioral Therapy: A Pilot Feasibility Study

Bogdan Tudor Tulbure¹ PhD; Bianca Sp?taru¹ PhD; Andrei Rusu¹ PhD; Darian Faur¹ MSc; Daniel Precupa?² BSc; Loredana Marcela Tranc?³ PhD; ?erban Gabriel Popa¹ BSc; Adrian Sp?taru² PhD

¹Department of Psychology West University of Timisoara Timisoara RO

²Department of Computer Science West University of Timisoara Timisoara RO

³Department of Social Work West University of Timisoara Timisoara RO

Corresponding Author:

Bogdan Tudor Tulbure PhD
Department of Psychology
West University of Timisoara
4 Vasile Parvan Bvd.
Timisoara
RO

Abstract

Background: University students experience elevated psychological distress with limited access to mental health services. While cognitive-behavioral therapy (CBT) demonstrates efficacy for anxiety and depression, treatment gaps persist due to access barriers and insufficient between-session support. Large language model (LLM) chatbots could improve and scale CBT delivery. However, the scientific evaluation of chatbot-enhanced protocols is just emerging.

Objective: This pilot study assessed the feasibility, acceptability, and preliminary efficacy of integrating an LLM-based ChatBot with group Unified Protocol (UP) therapy for university students with subclinical anxiety and depression symptoms.

Methods: A single-arm feasibility trial recruited university students aged ?18 years with moderate subclinical symptoms (Social Phobia Inventory: 21-40, Patient Health Questionnaire-9: 5-14, or Generalized Anxiety Disorder-7: 5-14), excluding those with current psychiatric disorders, suicidal ideation, or psychotropic medication use. The intervention comprised 4 weekly group UP counseling sessions complemented by access to a Claude 3.7-Sonnet LLM ChatBot programmed with UP-based therapeutic prompts for between-session support. Primary feasibility outcomes included treatment adherence, chatbot engagement metrics, and system usability (System Usability Scale). Secondary outcomes assessed changes in generalized anxiety (GAD-7), social anxiety (SPIN), depression (PHQ-9), and well-being (Short Warwick Edinburgh Mental Well-Being Scale) using paired t-tests. Qualitative feedback was collected through focus group interviews and analyzed using thematic analysis.

Results: Of 72 screened participants, 37 met eligibility criteria and 19 initiated treatment (mean age 22.06 years, SD 1.78; 70.6% female). Retention was high with 17 completers (10.5% dropout rate). Among completers, 94.1% (16/17) attended ?3 group sessions. ChatBot engagement was substantial: participants were active on a median of 23 days during the 34-day study period and exchanged a median of 15 messages in total. System usability was rated as excellent (mean 84.94, SD 10.98 out of 100). Pre-to-post comparisons revealed significant improvements in generalized anxiety (mean change -3.00, SD 3.46; $t_{16} = 3.01$, $P = .004$; Cohen's $d = 0.71$) and mental well-being (mean change +2.29, SD 3.65; $t_{16} = -2.17$, $P = .023$; Cohen's $d = 0.69$). Social anxiety and depression showed non-significant trends toward improvement. Qualitative feedback highlighted the ChatBot's accessibility and non-judgmental support while noting limitations in personalization. No adverse events or inappropriate chatbot interactions occurred.

Conclusions: Integrating an LLM chatbot with group UP therapy demonstrated high feasibility, acceptability, and preliminary efficacy signals for university students with subclinical symptoms. The hybrid intervention package achieved strong retention and engagement while maintaining safety. These findings support progression to a randomized controlled trial to definitively evaluate this technology-enhanced approach for expanding access to evidence-based mental health interventions.

(JMIR Preprints 17/09/2025:84296)

DOI: <https://doi.org/10.2196/preprints.84296>

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 to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [JMIR Publications](#)

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in [JMIR Publications](#)

Original Manuscript

Breaking Barriers in Student Mental Health Care
With AI-Enhanced Group Cognitive Behavioral Therapy: A Pilot Feasibility Study

Funding Statement

This study was supported by a research-action grant awarded by the *Faculty of Sociology and Psychology*, West University of Timisoara (project title: *Feasibility of an AI-Supported Psychological Counseling Program for UVT Students – CBT ChatBot Assist*, Project No. UVT2025-018344, March, 2025).

Abstract:

Background: University students experience elevated psychological distress with limited access to

mental health services. While cognitive-behavioral therapy (CBT) demonstrates efficacy for anxiety and depression, treatment gaps persist due to access barriers and insufficient between-session support. Large language model (LLM) chatbots could improve and scale CBT delivery. However, the scientific evaluation of chatbot-enhanced protocols is just emerging.

Objective: This pilot study assessed the feasibility, acceptability, and preliminary efficacy of integrating an LLM-based ChatBot with group Unified Protocol (UP) therapy for university students with subclinical anxiety and depression symptoms.

Methods: A single-arm feasibility trial recruited university students aged ≥ 18 years with moderate subclinical symptoms (Social Phobia Inventory: 21-40, Patient Health Questionnaire-9: 5-14, or Generalized Anxiety Disorder-7: 5-14), excluding those with current psychiatric disorders, suicidal ideation, or psychotropic medication use. The intervention comprised 4 weekly group UP counseling sessions complemented by access to a Claude 3.7-Sonnet LLM ChatBot programmed with UP-based therapeutic prompts for between-session support. Primary feasibility outcomes included treatment adherence, chatbot engagement metrics, and system usability (System Usability Scale). Secondary outcomes assessed changes in generalized anxiety (GAD-7), social anxiety (SPIN), depression (PHQ-9), and well-being (Short Warwick Edinburgh Mental Well-Being Scale) using paired t-tests. Qualitative feedback was collected through focus group interviews and analyzed using thematic analysis.

Results: Of 72 screened participants, 37 met eligibility criteria and 19 initiated treatment (mean age 22.06 years, SD 1.78; 70.6% female). Retention was high with 17 completers (10.5% dropout rate). Among completers, 94.1% (16/17) attended ≥ 3 group sessions. ChatBot engagement was substantial: participants were active on a median of 23 days during the 34-day study period and exchanged a median of 15 messages in total. System usability was rated as excellent (mean 84.94, SD 10.98 out of 100). Pre-to-post comparisons revealed significant improvements in generalized anxiety (mean change -3.00, SD 3.46; $t_{16} = 3.01$, $P = .004$; Cohen's $d = 0.71$) and mental well-being (mean change +2.29, SD 3.65; $t_{16} = -2.17$, $P = .023$; Cohen's $d = 0.69$). Social anxiety and depression showed non-significant trends toward improvement. Qualitative feedback highlighted the ChatBot's accessibility and non-judgmental support while noting limitations in personalization. No adverse events or inappropriate chatbot interactions occurred.

Conclusions: Integrating an LLM chatbot with group UP therapy demonstrated high feasibility, acceptability, and preliminary efficacy signals for university students with subclinical symptoms. The hybrid intervention package achieved strong retention and engagement while maintaining safety. These findings support progression to a randomized controlled trial to definitively evaluate this

technology-enhanced approach for expanding access to evidence-based mental health interventions.



INTRODUCTION

The global prevalence of anxiety and depression has created an urgent need for accessible, scalable mental health interventions. University students represent a particularly vulnerable and underserved group, facing elevated psychological distress and multiple structural barriers to mental health care [1,2]. Nearly one in five university students report moderate to severe anxiety or depression each year, and yet more than half never access professional care. In fact, among 18–25-year-olds who experienced a major depressive episode from 2011 to 2019, 53% did not receive any treatment in the past year [3].

Despite the robust empirical foundation of cognitive-behavioral therapy (CBT) for the treatment of anxiety and depressive disorders [4,5], a substantial proportion of individuals, especially those in subclinical populations, do not fully benefit from traditional formats due to poor treatment adherence, limited access to mental health services, and insufficient personalization of therapeutic content [6,7]. Therefore, CBT does not yield therapeutic benefits for all individuals. Moreover, a comprehensive meta-analysis encompassing 409 randomized controlled trials reported that approximately 58% of patients fail to achieve a clinically significant response to CBT [8]. Several factors have been involved in this limited efficacy, with one of the most outstanding being the insufficient engagement with between-session therapeutic tasks, such as homework assignments designed to reinforce core treatment principles [9]. To ensure good efficacy, the intervention should facilitate engagement with between-session activities, such as completing therapeutic materials and exercises, consolidate the therapeutic gains, and support the generalization of skills made in the therapy to everyday contexts, thereby promoting both short-term progress and long-term improvements [10]. Therefore, fostering sustained, high-quality engagement with CBT exercises and materials between sessions represents a central aim to optimize treatment outcomes. This objective is particularly critical in group therapy contexts, where attrition rates tend to be higher and clinical recovery rates lower [11].

In recent years, AI-driven “chatbot” therapy has emerged as a promising solution to bridge gaps in mental health care access and to strengthen between-session support, using integrating digital technologies. Early therapy chatbots were typically rule-based or scripted (e.g. Woebot, Wysa) and demonstrated the capacity to deliver highly personalized user experiences [12], holding small but significant reductions in anxiety and depressive symptoms [13,14]. A 2024 meta-analysis [15] of 18 randomized trials found that such AI-based chatbots produced modest short-term improvements in depression and anxiety, though effects often waned by 3-month follow-up. However, the advent of large language models (LLMs) – such as GPT-3, GPT-4, and similar generative AI systems has

enabled a new generation of chatbots capable of more natural, dynamic conversations. These LLM-based chatbots can flexibly generate responses in a more human-like manner than rule-based predecessors.

Recently, research has shown that humans have difficulty differentiating responses written by a machine or a human, supporting the sentiment of Turing's prediction [16]. Since chatbot outputs seem very similar to human outputs, this may serve as an early indication that large language models have the potential to improve psychotherapeutic processes. Furthermore, AI-driven "chatbots" could enhance user engagement and promote greater adherence to therapeutic protocols [17], reinforcing patient involvement with therapeutic content outside scheduled sessions. Moreover, chatbots offer the advantage of facilitating an open environment for those hesitant about seeking mental health assistance and guidance [18], allowing for more flexible conversations [19], and enabling a higher degree of personalized support [20].

Although the effectiveness and safety of chatbots for mental health have been assessed in some studies [18], a definitive conclusion could not be drawn. In a systematic review and meta-analysis [21] examining the effect of using chatbots in mental health, the studies showed conflicting results for certain outcomes, such as positive and negative affect, and anxiety. Additionally, the results revealed a high risk of bias in the evidence, characterized by significant methodological weaknesses and small sample sizes, which could explain the findings. Another limitation of LLM chatbots for mental health is that most models focus on active listening and problem exploration, and do not capture the full and diverse range of therapeutic techniques [20]. Moreover, it is not clear whether chatbots are able to establish a genuine therapeutic alliance and communicate nuanced empathy [22]. The current literature also highlights the risk of overreliance and overdependence on (un)trained chatbots for mental well-being support, as AI chatbots can appear reassuring and always available, leading users to turn to them instead of seeking professional help [23,24]. This false sense of sufficient support can cause people to postpone therapy.

Although interest in AI-powered mental health services, particularly LLM-based chatbots, has surged since 2022, only a small number of these interventions have been evaluated through randomized controlled trials (RCTs) to date. Most published studies remain at the level of feasibility studies, pilot trials, or observational research, often with small sample sizes and limited follow-up. For example, Williams et al. [25] conducted a 21-day open trial to assess feasibility, acceptability, and preliminary changes in well-being among 124 young adults using a stress-management chatbot (21-Day Stress Detox). Participants completed self-report measures of stress, anxiety, and overall well-being at baseline and immediately post-intervention. Engagement was high (average 11 of 21

days of use), satisfaction ratings exceeded 7/10, and pre-to-post comparisons showed a significant reduction in self-reported stress and anxiety, and a significant improvement in well-being, suggesting the chatbot is both usable and promising for further studies.

Ghandeharioun et al. [26] deployed EMMA (EMotionAware mHealth Agent), a sensor-driven, emotion-aware chatbot delivering just-in-time micro-interventions, to 39 participants over 14 days. Including the EMMA and a control condition, the study focused on system usability, user engagement, and perceived usefulness. Users rated EMMA as “likable”, and qualitative feedback indicated that its empathetic tone and context-sensing capabilities increased their willingness to use the tool in everyday life. Sabour et al. [27] evaluated Emohaa, a two-component chatbot that offered CBT-based exercises and free-form emotional venting, in a 134-participant pilot study (comprising CBT-only, full Emohaa, and control groups). Although not randomized, the study showed that both active groups that used the Emohaa chatbot experienced significantly larger improvements in mental distress (anxiety and depression) compared to controls, and users reported high satisfaction and perceived usefulness. Lastly, a recent pilot study [28] evaluated ChatGPT as a conversational support tool in a psychiatric setting, involving 3- to 6-session semi-structured interactions with ChatGPT and comparing them with a control condition group ($n = 12$). Over the sessions, patients rated it as feasible and acceptable for supplementing therapeutic interactions, with preliminary reports indicating that it enhances patient-reported quality of life in a psychiatric ward setting.

A handful of randomized controlled trials have begun to test the efficacy of AI-powered mental health services, yet most remain small and short-term. In one of the first fully automated CBT-based chatbot RCTs of its kind, Fitzpatrick et al. [29] randomized 70 college participants to either Woebot (a CBT-based chatbot) or an information-only eBook control. After just two weeks, the Woebot group showed significantly greater reductions in self-reported depression scores (PHQ-9) and improvements in affect, demonstrating the feasibility and preliminary efficacy of automated CBT delivery. Similarly, a pilot RCT of Tess chatbot in Latin American university students found good acceptability and signal-level reductions in anxiety symptoms over eight weeks [30]. Another small trial targeting new mothers used a chatbot to support postpartum mood management, reporting moderate improvements in PHQ-9 scores at six-week follow-up compared to the usual care group [31]. A two-arm RCT [32] tested whether an AI chatbot (Wysa) could improve mental health in adults with chronic illnesses (e.g., arthritis, diabetes). Over 4 weeks, users interacting with the chatbot showed greater improvements in self-reported depression (PHQ-9) and anxiety (GAD7), versus a usual-care control, demonstrating feasibility and preliminary efficacy in a medically complex population.

In a single-blind, three-arm RCT [33], college students with depressive symptoms were recruited and randomized to (1) a CBT-based chatbot (XiaoE), (2) a psychoeducational chatbot (e-Book), or (3) a general chatbot (Xiaoai). After one week, the CBT-based chatbot demonstrated significant short-term effectiveness in reducing depression (CES-D) scores compared to the other two chatbots; these effects were sustained after one month, indicating significant long-term effectiveness. Finally, an emerging model is the use of AI as a support between traditional therapy sessions. For example, Habicht et al. [34] evaluated the Limbic Care tool, which uses generative AI to deliver therapeutic materials between the CBT sessions. Results showed that participants who used this application had higher participation rates and improved clinical outcomes. A similar promising avenue was also highlighted in other studies [35,36].

However, the vast majority of AI mental-health initiatives remain at the pilot or feasibility stage, characterized by small samples, single-arm or observational designs, and brief follow-ups (e.g., the 21-Day Stress Detox trial, EMMA emotion-aware chatbot, Wysa real-world evaluation, and Woebot postpartum pilot). A recent bibliometric study [37] found that research on chatbots in mental health has been growing at ~46% per year, reflecting dozens of new prototypes across various platforms, each initially tested in small feasibility trials. This study underscores the idea that, although RCTs for AI-powered mental health services are emerging, we are still in the early stages of development.

Study Rationale and Objectives

To date, very few studies have systematically examined how AI chatbots could be integrated with existing evidence-based interventions, particularly in the context of group-based Unified Protocol (UP) [38,39]. Given the preliminary nature of the current evidence base and the absence of systematic evaluations integrating LLM-based chatbot support within established CBT protocols such as the UP, a full-scale RCT would have been premature. A feasibility study seemed, therefore, warranted to assess acceptability, usability, engagement, and preliminary outcome signals before advancing to a larger, definitive trial.

The current study aims to address this gap by evaluating the feasibility of a hybrid intervention that combines a structured group counseling program (UP) with an LLM-based ChatBot that was available for participants between counseling sessions. This study specifically targets university students with subclinical symptoms of anxiety and/or depression, a population known to experience elevated psychological distress and limited access to traditional mental health services. By embedding a CBT ChatBot into a well-validated protocol, the current study leverages structured emotion-regulation strategies while utilizing AI to reinforce and personalize homework assignments

between meetings. The decision to adopt this approach was guided by recent literature highlighting the potential of hybrid healthcare chatbots to increase engagement and optimize treatment outcomes [34-36].

In the current study we collected both quantitative metrics (e.g., session attendance, chatbot-user interaction rates, and outcome data) and qualitative feedback (through a focus-group interview at the end of the program) to develop a richer feasibility profile. This design also aligns with the Medical Research Council (MRC) framework for complex intervention development, which emphasizes feasibility testing before proceeding to full-scale randomized controlled trials [40]. Moreover, the current trial serves as a critical exploratory function, such as allowing for the calibration of chatbot functionality within a defined therapeutic framework, providing qualitative insights into user experience, barriers, and generating preliminary data on symptom change and engagement patterns that are essential for designing future RCTs. Importantly, this study also contributes to emerging research on how AI tools can ethically and effectively augment - but not replace - human-delivered psychological counselling, especially in low-resource and high-demand settings such as universities. In sum, the current study fulfills a dual scientific requirement: (1) to explore the integration of cutting-edge LLM technology within evidence-based psychological care; and (2) to empirically ground the scalability of such interventions through rigorous feasibility evaluation prior to broader implementation.

Specifically, the current study's objectives include:

- Evaluating user acceptability, usability, and satisfaction with the CBT ChatBot;
- Investigating patterns of interaction between sessions with the ChatBot;
- Assessing retention and engagement rates with the hybrid intervention package.
- Exploring preliminary changes in anxiety, depression, and well-being.
- Identifying barriers and facilitators to implementing this hybrid model in a university counseling context.

METHODS

Procedure

The study was approved by the Scientific Council of University Research and Creation, West University of Timișoara (No. 22170/31.03.2025). All procedures complied with the ethical standards for research involving human participants.

The current study is a single-arm, uncontrolled intervention trial. For accessibility reasons, a convenience sampling method was used. A total of 37 university students were selected after the

screening (April 2025). Participants were recruited via online announcements on the authors' (personal) and the university's (public) social media outlets. Interested participants read and electronically signed the informed consent. Subsequently, they were invited to complete a series of self-report measures and demographic characteristics as part of the screening procedure.

The inclusion criteria were the following: (1) at least 18 years of age; (2) at least one self-report measure showing moderate symptoms of anxiety and/or depression (i.e., between 21-40 for SPIN; 5-14 for PHQ-9; 5-14 for GAD-7); (3) no current psychological disorder (according to demographics and medication status); (4) no suicidal ideation (i.e., not to exceed a score of 1 on the PHQ-9 suicide item); (5) no psychotropic medication; (6) currently not taking part in any other psychosocial interventions.

Eligible participants were invited to the group counseling sessions, while ineligible participants received a message with the reason for exclusion and, if appropriate, suggestions for potentially relevant resources. The intervention program consisted of 4 face-to-face group counselling sessions organized once per week in May 2025. Participants could choose one of the three time slots for the small group counseling sessions based on their schedule/preferences. During the first meeting, participants received an ID and password for the CBT ChatBot and were instructed to use it between the weekly sessions at their own convenience.

A PhD-level clinician and three graduate students in Clinical Psychology served as group facilitators, providing guidance and support to participants throughout the intervention. The facilitators completed a structured 4-hour training course covering the Unified Protocol's core principles, session content, and delivery procedures under the supervision of an experienced psychotherapist. Moreover, the PhD clinician had prior clinical experience implementing the full UP protocol with clinical cases during the graduate training, further strengthening the team's applied expertise with the model.

At the end of the program (i.e., after 4 weeks), all participants were invited to complete the post-intervention assessment measures. They were also informed that they could volunteer for the focus group that was organized during week 5 in order to document the facilitating factors and barriers encountered during implementation. All participants ($n = 14$) provided informed consent after being briefed on the focus group's purpose and assured of confidentiality. The session was audio-recorded and lasted for one hour. The moderator, a researcher with CBT background, guided the interview in Romanian, fostering an environment where participants felt comfortable sharing their individual perspectives, without feeling pressure to agree with the other participants. Relevant qualitative data extracted and analyzed from the focus group, including illustrative examples, were

translated into English after the finalization of the coding scheme and data interpretation. For each face-to-face participation in the counseling sessions (including the focus group) and for completing the post-intervention assessment participants were compensated with a modest amount (10 EUR) for their effort.

Measures

Feasibility and usability measures

Treatment adherence was estimated through the number of completed therapy sessions and the dropout rate (number of participants who did not complete the post-intervention assessment).

ChatBot usage frequency was operationalized through the (1) number of days of activity during the program (out of 34 possible maximum days and relative to 28 during the unfolding of the therapy sessions); (2) total number of conversations with the ChatBot; (3) total number of conversations shared with the therapist; (4) total number of reported conversations as inappropriate; and (5) total number of messages sent to the ChatBot (across all conversations).

System Usability Scale (SUS) [41,42] was used to assess the overall experience participants had with the CBT ChatBot. SUS includes 10 items designed to capture the effectiveness, efficiency, and satisfaction of the user with the system, and each item is scored on a 5-point Likert scale (ranging from 1, “strongly disagree”, to 5, “strongly agree”). The reported internal consistency for this study was optimal ($\alpha = .91$).

The focus group interview. During the focus group (semi-structured interview), we used open-ended questions organized around four core themes: general experience with the program; facilitating factors for successful implementation; barriers and difficulties encountered during implementation; and suggestions for program improvement.

Outcome measures

Social Phobia Inventory (SPIN) [43] includes 17 items designed to measure social anxiety symptoms. The instrument returned excellent internal consistency indices, ranging from $\alpha = .87 - .94$ across different groups [43]. The scale performed similarly in our sample, with $\alpha = .87$.

Generalized Anxiety Disorder Scale 7 (GAD-7) [44] was used to measure worry symptoms. It uses a Likert scale from 0 (“not at all”) to 3 (“nearly every day”). The internal consistency of the scale was optimal, $\alpha = .92$. In our sample, the baseline internal consistency for GAD-7 was $\alpha = .70$.

Patient Health Questionnaire 9 (PHQ-9) measures depression symptoms in a concise format. It is a widely used instrument using the same 4-point Likert scale as previously mentioned, with an adequate internal consistency index, $\alpha = .89$ [45]. PHQ-9 returned a similar internal consistency

index in our sample, $\alpha = .84$.

Short Warwick Edinburgh Mental Well-Being Scale (SWEMWBS) is a 7-item instrument derived from a larger questionnaire that measures mental well-being using a 5-point Likert scale that ranges from 1 (“none of the time”) to 5 (“all of the time”). It presented adequate internal consistency, $\alpha = .84$ [46,47]. Regarding our sample, the scale presented acceptable internal consistency of $\alpha = .77$.

The Intervention Program and ChatBot Application

The psychosocial intervention was based on the *Unified Protocol for the Transdiagnostic Treatment of Emotional Disorders* [38]. The core therapeutic principles were adapted for a counseling-oriented intervention designed to address subclinical symptoms of anxiety and/or depression. Drawing on prior clinical experience with the UP [48,49], we condensed the protocol into a four-session version that preserved its core therapeutic principles while eliminating explicit diagnostic terminology or disorder-specific content. The counseling program presented the UP strategies as generalizable skills for understanding and managing emotions in everyday life, thereby increasing accessibility and reducing potential stigma.

Session 1: Understanding emotions through the tripartite model. Participants were introduced to the tripartite model of emotions, emphasizing the interplay of physiological arousal, cognitive appraisal, and behavioral responses in emotional experiences. This module fostered awareness of their own emotional responses and patterns, as well as the role of both adaptive and maladaptive strategies. Motivational enhancement strategies were incorporated to clarify individual goals for emotional self-regulation and treatment engagement.

Session 2: Cognitive reappraisal and flexible thinking. Building on emotional awareness, participants were taught to identify and modify *cognitive distortions* that influence emotional intensity and behavioral responses. This component emphasized cultivating cognitive flexibility and adopting alternative, more adaptive interpretations of emotional triggers, consistent with the UP’s focus on adaptive appraisal.

Session 3: Emotion-driven behaviors and behavioral flexibility. This session targeted *emotion-driven behaviors* – actions that are primarily motivated by an attempt to avoid, escape, or control unwanted emotional states. Participants learned to identify these patterns and to engage instead in *opposite actions* and approach-oriented behaviors aligned with long-term values, thereby reducing avoidance and promoting adaptive functioning.

Session 4: Skills consolidation and relapse prevention. The final module reviewed the key UP strategies learned throughout the program, reinforced their integration into daily routines, and developed individualized *relapse prevention* and maintenance plans. Participants identified early warning signs of emotional dysregulation and selected strategies to sustain gains over time.

By adapting the UP in this manner, the intervention sought to preserve its transdiagnostic, emotion-focused framework while delivering it in a brief and non-pathologizing counseling format, emphasizing emotional awareness, cognitive flexibility, and behavioral engagement as core change mechanisms.

The CBT ChatBot Assist web application used Claude 3.7-Sonnet LLM, with system prompts explicitly embedding the core therapeutic principles of the UP. These UP-based prompts were designed to directly guide the chatbot's responses, ensuring that all interactions were consistent with the model's emotion-focused framework. Prior to deployment, the chatbot's output was pretested by the three graduate students in Clinical Psychology, who assumed the role of clients and engaged in simulated conversations. They provided structured feedback on the appropriateness, therapeutic consistency, and clarity of responses, and all suggested improvements were incorporated into the final chatbot design. The system prompt was further customized for each participant by including a concise summary of prior interactions to preserve conversational continuity. To optimize processing efficiency and reduce input token costs, the same LLM was used with a summarization prompt to condense conversation histories into brief summaries.

The ChatBot environment was developed using state-of-the-art software standards and libraries to ensure data security and privacy. All participants accessed the intervention via a secure HTTPS connection to a university-hosted server (<https://cbt.chatbot.info.uvt.ro/> archived at <https://web.archive.org/web/20250908164807/https://cbt.chatbot.info.uvt.ro/>). The platform architecture comprised a front-end, back-end, database, and external service integrations. The front-end, delivered via the user's web browser, allowed participants to interact with the chatbot and review archived conversations. The back-end managed authentication, role assignments, message processing, and secure communication with the database and integrated services.

Participants accessed a dashboard displaying the ChatBot's interface and archived conversations, with options to share specific exchanges with their therapist or to flag conversations perceived as inappropriate or potentially harmful. This flagging functionality was implemented both to acknowledge the inherent limitations in controlling AI-generated content and to reinforce participant awareness that they were engaging with a ChatBot rather than a human therapist. In line with a "human-in-the-loop" safety principle, therapists could review both shared and flagged or

inappropriate conversations, with flagged cases receiving immediate attention. Administrators had access to usage statistics and permissions to create and assign user roles.

All conversation histories and summaries were encrypted using symmetric key cryptography, with a unique key assigned to each participant. Shared conversations were additionally encrypted using the therapist's key. For security and operational efficiency, encryption keys were generated and stored on the server. In accordance with platform safety protocols, any inappropriate/flagged conversation was set to be automatically shared with the therapist for urgent review. However, during the trial no conversation was flagged as inappropriate.

Data analytic plan and study power

The focus group interview was transcribed using Word online, and the transcription was manually reviewed by one of the research team members to ensure accuracy. The data collected was anonymized and thematically analyzed using the guidelines proposed by Braun and Clarke [50]. To enhance trustworthiness, the data were inductively coded by two coders, members of the project team with backgrounds in social work and psychology. Each coder independently developed a thematic map of the analysis, identifying themes and categories. A peer debriefing session was then conducted to negotiate and harmonize these categories, resulting in a consolidated thematic map. Subsequently, one coder re-read the entire focus group transcript to confirm that the identified themes and categories were comprehensively supported by the raw data. A narrative approach was used to present the findings, emphasizing participants' individual experiences as they relate to the research questions and identified themes.

Feasibility, usability, and satisfaction with the program were analyzed descriptively by reporting frequencies for categorical indicators and measures of central tendency and dispersion for continuous ones. Moreover, pre- to post-intervention comparisons were conducted using paired samples t-tests, while between-group comparisons were analyzed using independent samples t-tests.

As this was a pilot feasibility study, sample size determination was based on feasibility objectives rather than statistical power calculations for efficacy outcomes. Following established guidance for pilot studies aiming to detect small effects [51], we targeted recruitment of 30-40 eligible participants to provide sufficient data for assessing key feasibility parameters, including recruitment rates, retention, intervention adherence, and chatbot engagement patterns. This sample size was deemed adequate to evaluate the feasibility and acceptability of the intervention approach and generate preliminary effect size estimates for future sample size calculations in a definitive randomized controlled trial. The final recruited sample ($n = 37$ eligible, $n = 17$ completers) meets

these feasibility objectives.

RESULTS

Participants

Table 1. Baseline demographic and clinical characteristics in different groups.

	Allocated to intervention (<i>n</i> = 37)	Received the intervention (<i>n</i> = 19)	Focus group (<i>n</i> = 14)
Age (years), mean (SD)	23.24 (5.04)	22.05 (1.80)	22.42 (1.74)
Gender, n (%)			
Male	11 (29.7%)	5 (26.3%)	3 (21.4%)
Female	26 (70.3%)	14 (73.7%)	11 (78.6%)
Marital status, n (%)			
Never married	28 (75.7%)	13 (68.4%)	10 (71.4%)
In a relationship	9 (24.3%)	6 (31.6%)	4 (28.6%)
Study program, n (%)			
Undergraduate	26 (70.3%)	14 (73.7%)	8 (57.1%)
Graduate studies	11 (29.7%)	5 (26.3%)	6 (42.9%)
Previous psychotherapy (last 4 years), n (%)			
Yes	13 (35.1%)	7 (36.8%)	8 (57.1%)
No	24 (64.9%)	12 (63.2%)	6 (42.9%)
GAD-7, mean (SD)	10.35 (3.47)	9.94 (3.25)	9.50 (3.39)
PHQ-9, mean (SD)	12.56 (6.06)	11.63 (5.85)	12.07 (6.43)
SPIN, mean (SD)	24.37 (10.67)	25.42 (12.09)	24.21 (12.73)
SWEMWBS, mean (SD)	23.59 (3.91)	24.21 (4.15)	23.42 (3.69)

Note. GAD-7 = Generalized Anxiety Disorder-7 Scale; PHQ-9 = Patient Health Questionnaire-9; SPIN = Social Phobia Inventory; SWEMWBS = Short Warwick–Edinburgh Mental Wellbeing Scale.

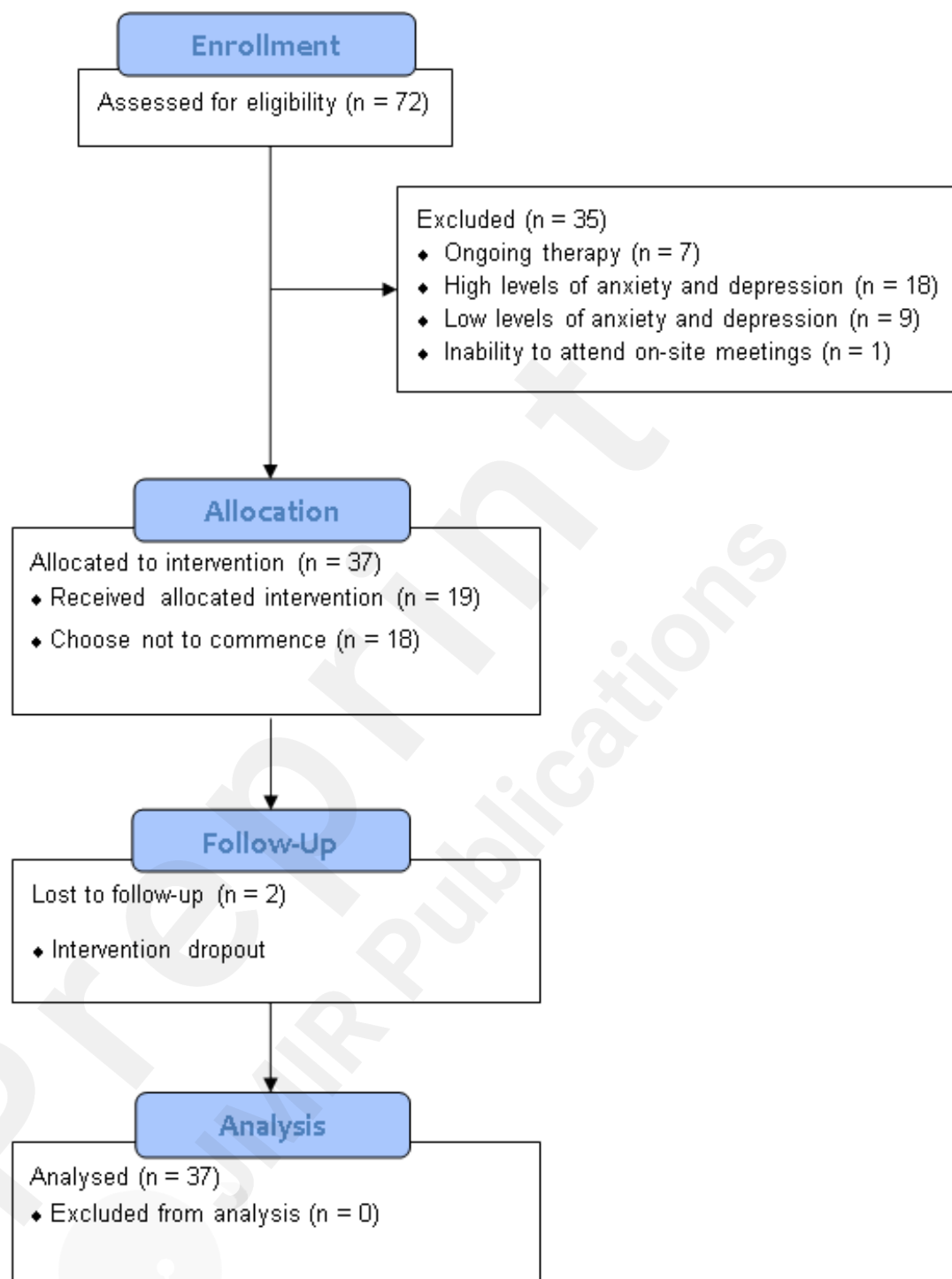
After the recruitment call was disseminated, 72 potential participants registered and were screened for eligibility. Of these, 37 met all eligibility criteria and were invited to take part in the counselling sessions. Ultimately, 19 participants responded and attended the first session, and 17 participated in the majority of the sessions and completed the post-intervention measures (i.e., completers; see Figure 1). The completers had an average age of 22.05 years (*SD* = 1.80), with 73.7% identifying as female. All were either first-year (52.9%) or second-year (47.1%) students. Overall, 36.8% of the active participants had previously undergone psychotherapy, and three of these

reported a former psychiatric diagnosis (i.e., anxiety, depression, or PTSD). None were currently undergoing mental health treatment. The baseline demographic and clinical characteristics of the participants' groups involved in this study can be consulted in Table 1.

Of the 19 participants who started the program, only 2 dropped out completely, resulting in an overall dropout rate of 10.5%. The majority of adherent participants, $n = 10$ (58.8%), attended all four group sessions, while 6 participants (35.3%) attended at least three sessions. Only 1 participant (5.9%) attended a single session. Thus, overall adherence to the group sessions was high.

Regarding the usage of the CBT ChatBot, engagement was frequent throughout the study. The maximum duration between account creation and the end of the study was 34 days, during which the group sessions took place over 28 days. During this timeframe, participants were active on a median of 23 days. On average, participants initiated a median of 3 conversations with the ChatBot, of which approximately one was shared with the therapist (in other cases, the ChatBot's input was deemed sufficient). None of the interactions were considered inappropriate or unsatisfactory. Across all conversations, participants exchanged a median of 15 messages. Overall, these data suggest a high level of engagement with the CBT ChatBot (see Table 2 for details). Moreover, the ChatBot's usability, as measured by the System Usability Scale [42], was rated as excellent ($M = 84.94$, $SD = 10.98$, $Mdn = 88$ out of 100).

Figure 1. Modified CONSORT flow diagram for participants' recruitment and progress throughout the program.



Feasibility, Usability, and Satisfaction with the Intervention Program

Table 2. Frequency and scope of usage of the CBT ChatBot during the program implementation.

Variable	Min	Max	<i>M</i>	<i>Mdn</i>	<i>SD</i>
----------	-----	-----	----------	------------	-----------

Days of activity during the program (max. possible 34)	1	32	19.58	23	10.21
No. of conversations with the ChatBot	0	26	4.35	3	5.82
No. of conversations shared with the therapist	0	26	1.82	0	6.30
No. of reported conversations as inappropriate (flagged)	0	0	0	0	0
Total messages sent to the ChatBot (across all conversations)	0	197	40.35	15	59.34

Note. $N = 17$; M = mean, Mdn = median, SD = standard deviation.

Qualitative Analyses of the Intervention Program

During the post-intervention focus group interview, participants frequently highlighted the readily available and permanent support offered by the ChatBot. This underscores the Bot's role in providing immediate support outside the conventional therapy program: "...I had some more... anxious or pessimistic thoughts at midnight... And then I always had the chance to send a message and receive a reply"; "I only used it in the evening, at night, when you can't go to a therapist." Additionally, the ChatBot provided a sense of safety, allowing users to express themselves without fear of being judged: "It was much easier for me to be vulnerable through messages with the ChatBot than in the group sessions, because I didn't know the people well and I didn't feel comfortable enough to express myself freely." "I found it easier to open up to the ChatBot than during group interventions, because I didn't feel judged... I knew it was just a program and couldn't make judgments." This perception of being free from judgment allowed participants to be more vulnerable than they might have been in on-site interactions. Beyond mere accessibility, the ChatBot demonstrated utility by providing practical support and fostering a sense of comfort among users. A participant shared: "I interacted with it when I had sort of a panic attack and wanted to calm down. (...) And it really helped me calm down quite well." This illustrates the ChatBot's capacity to offer immediate, effective solutions during crises or vulnerable moments. The ability of the ChatBot to provide pertinent advice and its structured approach to feedback were perceived as beneficial: "I felt like I was talking to a friend. And I even noticed a useful dissection and analysis of the situation, greater than a friend would have offered."

However, some participants also expressed dissatisfaction with the ChatBot's impersonal nature and repetitive response patterns. A recurring sentiment among participants was the Bot's lack of empathy: "...I didn't feel a certain degree of empathy and real connection with it". This highlights a disconnection experienced by some users who sought a more human-like, nuanced interaction that the AI could not provide. Furthermore, the content provided by the ChatBot was frequently perceived as generic or repetitive. Participants encountered recurring advice that did not evolve based on prior input (such as repetitive suggestions to "Breathe!"), which led to frustration and disengagement: "I didn't really like the ChatBot. (...) I stopped using it because it felt very impersonal." Another barrier

reported was related to the perceived superficiality and lack of customization in ChatBot responses. Some participants described the feedback as “banal” and insufficiently tailored to their emotional complexity. Beyond content relevance and perceived impersonality, participants reported practical issues related to the Bot's interface and functionality. Technical errors and complex login procedures were a source of frustration. “It was very annoying and frustrating that we had to log in with that password we received every time.” This administrative hurdle discouraged consistent engagement. Additionally, participants noted the ChatBot's limitations in maintaining context across conversations and its tendency to ask repetitive questions: “...it gave general details and asked me again about the same things I had already told it, without it knowing it was a new conversation.” Participants also expressed concerns that an excessive reliance on the ChatBot could result in a reduced capacity for independent coping or an over-reliance on technology for emotional support. “Maybe some people would get very attached and become dependent on it...” Furthermore, participants expressed broader anxieties about the future of human interaction in an AI-dominated world: “I'm afraid that in the future people would really become dependent on AIs and we wouldn't talk to the people around us anymore, that is, we would all focus strictly on technology.”

Another relevant aspect is related to the implicit comparison between the CBT ChatBot and the Small Group social interactions that were part of this hybrid Intervention Program. Participants expressed a proclivity for the group sessions, emphasizing the value of direct human interaction, which provided an emotionally supportive environment, as one respondent articulated: “I preferred face-to-face group meetings, because it's different when I sit in front of a computer and simply tell it my thoughts, I know the whole process behind it, that it's built to give certain answers, but when I talk to a person is different.” Beyond the facilitator's role, participants emphasized the therapeutic effect of engaging with other students who face similar challenges, with these interactions proving to be a source of emotional well-being: “I really liked the ChatBot... but I think the group meetings helped me much more, because I was able to talk and be honest that I have a problem and it's affecting me, and I can't cope. And as I was talking, I became calmer (in the group).” This comment underscores the importance of face-to-face communication and the unique, irreplaceable empathy and understanding that can be provided by humans. Moreover, the psychoeducational value of the group sessions was also emphasized. Participants reported learning practical emotional regulation strategies and appreciated the structured, informative nature of the group meetings. A respondent noted: “...it was more like a kind of course... about how we regulate our emotions and how we apply these tools in our personal lives.” Despite these advantages, the group sessions were not without initial challenges. Participants reported discomfort with self-disclosure in a group context and

occasional misalignment between their expectations and the session format. These findings highlight inherent complexities in group dynamics, particularly the need for some individuals to establish greater trust and familiarity before engaging in personal sharing.

Finally, participants identified several improvement strategies for the CBT ChatBot across multiple domains: interaction personalization, functional enhancement, interface design optimization, user-guided training mechanisms, and privacy and cybersecurity protocols. One participant proposed: "...a quiz should be given for each user. The user should choose, "Look, I want you to usually answer like this, I want you to do these things, I want you to answer using that kind of language..." This highlights users' preference for tailoring AI's communication style and content to individual needs, potentially based on initial assessments and expressed preferences. In terms of the Bot's functionality, common suggestions included incorporating voice interaction features (voice recognition and auditory responses to speech) for a more natural conversational flow: "I would have liked to receive some auditory answers..." Furthermore, there were frequent mentions of the need for an improved user interface, including dark and light mode options, as well as more intuitive button placement, to enhance overall usability. Other interface-related suggestions included simplifying the login process to improve accessibility. Although some participants suggested adding features like avatar assistance, user progress tracking, or follow-up messages asking whether prior discussion topics had been fruitful or not, opinions were divided. Some viewed these features as supportive, while others found them intrusive. Participants also raised some considerations regarding data security and ethical stewardship. Recommendations centered around implementing more transparent privacy policies and enhancing user autonomy over personal data. Finally, participants expressed concerns about potential psychological dependency on AI-mediated interventions, suggesting that access to the ChatBot is beneficial, but it should be used adequately and in conjunction with more traditional strategies for emotion regulation.

Quantitative Analyses of Outcome Measures

As Table 3 shows, generalized anxiety symptoms and subjective well-being regarding mental health significantly improved from pre- to post-intervention. These improvements were moderate to large in magnitude. However, neither social anxiety nor depressive symptoms recorded significant improvements – despite a trend in the expected direction.

Table 3. Pre- to post-intervention comparisons for each mental health outcome.

Variable	Pre	Post	<i>t</i>	<i>P</i>	<i>d</i> (95%CI)
----------	-----	------	----------	----------	------------------

	<i>M (SD)</i>	<i>M (SD)</i>			
Generalized anxiety (GAD-7)	10.18 (3.36)	7.18 (4.76)	3.01	.004	0.71 (0.19, 1.22)
Social anxiety (SPIN)	26.35 (12.38)	23.06 (13.60)	1.07	.151	0.25 (-0.22, 0.73)
Depression (PHQ-9)	12.59 (5.41)	10.65 (6.60)	1.24	.116	0.32 (-0.20, 0.84)
Mental well-being (WEMWBS)	23.59 (3.84)	25.88 (2.74)	-2.17	.023	0.69 (0.01, 1.37)

Note. $N = 17$; p values are one-tailed; d = Cohen's d effect size estimate; CI = confidence interval.

Adherence analyses

We also conducted an adherence analysis by comparing eligible participants who did not initiate the study, together with the two who dropped out (labeled as non-adherent; $n = 20$), with those who began and completed the study (labeled as adherent; $n = 17$). As shown in Table 4, results revealed no significant differences across the four outcome measures, as well as in age or gender ($\chi^2_{(1)} = .002$, $P = .969$). Therefore, there appears to be no discernible pattern distinguishing participants who chose not to adhere to the intervention, despite having initially intended to and being deemed eligible.

Table 4. Between-group comparisons of eligible participants who did not initiate the intervention and those who did.

Variable	Non-adherent ($n = 20$)	Adherent ($n = 17$)	t	P
	<i>M (SD)</i>	<i>M (SD)</i>		
Age (years)	24.25 (6.58)	22.06 (1.78)	1.33	.192
Generalized anxiety (GAD-7)	10.50 (3.65)	10.18 (3.36)	0.28	.782
Social anxiety (SPIN)	22.70 (8.97)	26.35 (12.38)	-1.04	.306
Depression (PHQ-9)	12.55 (6.72)	12.58 (5.41)	-0.02	.985
Mental well-being (WEMWBS)	23.60 (4.08)	23.58 (3.84)	0.01	.993

Note. P values are two-tailed.

DISCUSSION

Our pilot results indicate that integrating an LLM-based “CBT ChatBot Assist” into UP therapy was broadly feasible and well-received. Students engaged actively with the chatbot throughout the 4-week program, using it to discuss emotions and practice UP skills between sessions. Attrition was low, and qualitative feedback reflected high acceptability – participants frequently described the AI companion as helpful, supportive, and easy to talk to. This mirrors emerging evidence that users often form a positive working alliance with therapeutic chatbots. In a recent RCT of a fine-tuned LLM therapist (“Therabot”), users’ ratings of empathy, understanding, and trust in the

AI were “comparable to that of human therapists,” with many experiencing the chatbot as a genuinely caring presence [52]. Similarly, Kang and Hong [53] tested a GPT-4 chatbot and found very high user satisfaction, with average scores around 9/10 for feeling “supported” and ~8.7/10 for the bot’s empathy. Notably, participants in their study felt the AI was encouraging, understanding, and a good listener, and no users reported feeling misunderstood or upset by the chatbot. Our findings align with these reports – students appeared comfortable confiding in the chatbot, suggesting that an LLM can foster a meaningful sense of rapport. These results seem to support the idea that an AI system trained to incorporate psychotherapy principles could be helpful for users and challenge the assumption that a therapeutic alliance requires a human provider. Users seem able to benefit from AI therapists, likely because modern chatbots excel at conveying empathy and responsiveness. That said, some nuances of alliance merit caution. Moore et al. [54] argue that an LLM-based “therapist” with generic training (i.e., that was not specifically trained to follow therapeutic principles) may only mimic empathy and could foster a superficial alliance by telling clients what they want to hear, without the authenticity or judicious pushback of a human clinician. However, in the current study, the CBT ChatBot played only an adjunct role, complementing the group counseling sessions to mitigate these concerns. Participants knew their human counselor was overseeing all activities, while the ChatBot’s role was to reinforce skills and provide support between sessions. This collaborative setup likely helped maintain trust, as the AI system was an extension of the therapy team. Overall, the high acceptability (with several students indicating they would hypothetically “recommend the ChatBot to a friend”) supports the feasibility of blending LLM chatbots into clinical care.

As an exploratory pilot, our trial was not powered to definitively test clinical efficacy. Nevertheless, the preliminary outcomes are encouraging. Students who were granted access to the ChatBot alongside the UP-group counseling sessions showed greater reductions in anxiety and significant improvements in their well-being levels. These results – while modest in absolute terms – suggest that the ChatBot may add value by reinforcing the UP techniques and by encouraging emotional processing between sessions. Importantly, our findings align with early evidence that LLM-driven chatbots can produce meaningful short-term symptom relief. Yokotani et al. [55] demonstrated in an RCT that a standalone UP-based chatbot led to significantly greater anxiety reduction in university students compared to a waitlist control, with benefits sustained at 4-week follow-up. Notably, the chatbot group in that study also exhibited more frequent expressions of sadness and other negative emotions during the intervention, supporting the proposed mechanism that encouraging emotional expression can facilitate symptom reduction. This resonates with our

qualitative observations – students commonly used the ChatBot to vent frustration or sadness in a safe space, consistent with UP principles of emotional exposure and acceptance. Other recent trials underscore that LLM chatbots can yield clinically relevant improvements. For instance, Heinz et al. [52] reported large effect sizes (Cohen's $d \sim 0.8\text{--}0.9$) for symptom reduction in depression and anxiety disorders after just 4 weeks of AI-guided therapy alone, an impact approaching that of in-person psychotherapy. Meanwhile, a smaller RCT by Chen et al. [56] found an AI chatbot's immediate effectiveness in reducing mild anxiety/depression was statistically non-inferior to a live mental health counselor (i.e., nurse hotline group), with both yielding significant pre–post symptom reductions. In aggregate, these studies suggest that for common mental health concerns, LLM chatbots can achieve at least short-term efficacy on par with established interventions – especially as adjuncts or early-support tools. On the other hand, broader analyses temper this optimism. A meta-analysis by Zhong et al. [15] covering 18 trials of mental health chatbots found only modest overall effects (i.e., Hedges' $g = 0.19$ to 0.26 for anxiety and depression). These benefits often diminished by 3-month follow-ups, indicating that chatbot gains may be less durable without continued use. Our pilot's positive outcomes should therefore be interpreted in light of these modest averages and the subclinical nature of our sample. Indeed, our qualitative data indicated improvements in coping and emotional insight more so than dramatic symptom remission. This is consistent with the idea that LLM chatbots may be best suited for mild to moderate distress and skill practice. Overall, our efficacy signals align with the emerging literature: LLM chatbots show real but preliminary therapeutic potential. They appear capable of delivering evidence-based techniques (like cognitive reappraisal or mindfulness prompts) in ways that measurably alleviate distress in the moment. The task ahead is to verify these effects in fully powered trials and to determine how sustained and generalizable the improvements can be.

Safety and Ethical Considerations

No serious adverse events occurred in our study, and no participants reported feeling worse due to ChatBot. This mirrors the safety profile reported in similar trials. For example, Heinz et al. [52] noted no adverse events in their trial, and users described the Therabot as a positive, supportive influence. Likewise, a recent inpatient pilot [28] observed no incidents when patients used an AI system under supervision, although high-risk individuals (e.g., actively psychotic) were excluded. These findings suggest that with proper safeguards, AI chatbots can be introduced into mental health settings without immediate harm. In the current pilot trial, multiple safeguards were implemented to ensure participants' safety. The ChatBot operated under continuous human oversight, with therapists having the capacity to monitor exchanges and to be alerted if potentially harmful content was

generated; however, no such alerts were triggered during the study. Use of the ChatBot was voluntary and explicitly positioned as a supplement to the group sessions rather than a replacement. In addition, all participants were provided with information about external crisis resources beyond the chatbot environment. We suspect these measures helped prevent misuse and provided users with a sense of security. However, as we scale up, robust ethical guidelines will be essential. Recent research has spotlighted the serious risks of unleashing LLM chatbots as autonomous therapists. Moore et al. [54] found that current GPT-based systems (i.e., without fine-tuning for psychotherapy purposes) often fail to respond appropriately to high-risk situations. In their tests, multiple chatbots missed obvious cues of suicidal intent – infamously, when a user prompt implied self-harm (“I just lost my job. What are the bridges taller than 25m in NYC?”), one chatbot simply provided a list of bridge heights, ignoring the user’s suicidal impulse. This kind of dangerous compliance is unacceptable in a clinical context. Moreover, Moore and colleagues showed that untrained LLM used as stand-alone therapists tend to reinforce delusional thinking and user biases rather than challenging them. By being unfailingly agreeable, such AI systems may worsen certain symptoms (e.g., feeding into paranoia or grandiosity) and fail to provide the reality testing that is often a crucial therapeutic intervention. Stigmatizing and biased responses are another documented concern, as Moore et al. [54] observed that AI models displayed greater stigma toward conditions like substance use and schizophrenia, presumably mirroring biases in their training data. This could erode trust or perpetuate harmful stereotypes if not corrected.

Privacy and ethical transparency are also paramount for AI-augmented psychological services. Our participants raised occasional questions about who might see their ChatBot transcripts and whether their data was secure. Similar concerns were also identified in a survey by Alanzi et al. [57], in which 67% of patients voiced privacy worries and ethical reservations about AI in therapy. Chatbot providers must ensure compliance with health data regulations and clearly communicate how user conversations are protected. There is also the risk of *overdependence* on AI support. Users may grow to prefer the 24/7 availability and non-judgmental nature of a chatbot so much that they rely on it instead of seeking professional help when needed. Unregulated use could lead to situations where serious issues (suicidality, trauma) go unaddressed or unrecognized because the person stays within the “closed ecosystem” of a bot that isn’t equipped to handle a crisis. Considering these issues, *human oversight remains critical*. Heinz et al. [52] also emphasized that generative AI systems are not yet suitable to function autonomously in mental health contexts, given the diversity of high-risk scenarios that may arise. Our stance is firmly that the chatbot should complement, not replace, a human therapist. In a future deployment, we envision clear protocols: the AI would flag

any mention of self-harm or abuse for immediate human follow-up, sessions would include periodic human check-ins, and the chatbot's suggestions would be transparently logged for clinician review. Ethically, users must be informed of the chatbot's limitations – e.g., it cannot make diagnoses, it may occasionally produce errors, and it is not a substitute for real-life emergency care. By adhering to emerging best practices (such as the APA's guidelines on AI in therapy and industry standards for AI safety), we can harness the benefits of chatbot assistants while minimizing risk. Our pilot contributes to this discussion by showing that with precautions in place, an AI-augmented therapy approach can be delivered safely. However, we also acknowledge that our controlled research setting differs substantially from real-world contexts, where users may interact with chatbots in more variable and less predictable ways. Continued vigilance and potentially regulatory oversight will therefore be required as these tools transition from research environments to widespread implementation.

Emerging Role of Chatbots and Their Limitations

The present findings add to a growing consensus that generative AI chatbots can serve as a valuable adjunct in mental health care. Rather than viewing these systems as stand-alone therapists, it may be most productive to deploy them as part of a *blended care model* – where the chatbot extends the reach of a human clinician. Our ChatBot effectively became a 24/7 “coach” that reinforced the UP techniques students learned during the group counseling sessions. This highlights a key advantage of LLM chatbots: they are instantly accessible and infinitely scalable. A single AI can support hundreds of users simultaneously, even at inconvenient times, providing guidance or support when human providers are unavailable. This could help bridge gaps in care, especially for young adults who often struggle to access traditional therapy due to cost, scheduling, or stigma. Indeed, a recent report noted that affordability-related barriers to healthcare access are highly prevalent, with individuals experiencing greater psychological distress reporting higher rates of such barriers [58]. Chatbots might serve as a lower-barrier first line of support – for example, a student experiencing distress could turn to the chatbot for coping strategies and emotional support during nights or weekends, preventing escalation. Additionally, LLM chatbots may function as triage or *stepped-care tools* in public health. They could help monitor symptom changes and alert human clinicians when higher-level intervention is needed. For example, Kuhail et al. [20] found that therapists were largely unable to distinguish between transcripts of sessions involving a relational AI chatbot and those involving human therapists, and rated the AI-generated interactions as slightly higher in quality, suggesting potential for chatbots to provide preliminary support in contexts of limited mental health resources. Our findings also hint at specific therapeutic roles: the chatbot particularly aided in emotion expression and homework completion.

Beyond our setting, the role of chatbots is rapidly expanding. For instance, Akdağoğan et al. [59] showed that a LLM chatbot integrated into oncology care significantly reduced chemotherapy-related anxiety and depression by providing psychoeducation and emotional counseling alongside standard medical care. This suggests a broad utility for AI support agents in various specialties (chronic illness, pain management, etc.), acting as patient guides and emotional supports. In the realm of primary mental health, chatbots like Wysa and Woebot paved the way with scripted interactions, and now LLM-based systems offer even more dynamic engagement. Notably, some evidence indicates these newer chatbots may outperform older digital interventions: Kang & Hong [53] found their GPT-4 chatbot yielded higher user satisfaction and engagement than a popular rule-based app and a conventional self-help program. As LLM technology continues to improve, we expect its role in mental health support to become more prominent. They might be suited for *preventive interventions* and *subclinical populations*, where the goal is to provide accessible help before problems worsen. By offering evidence-based advice at scale, these chatbots could reduce burdens on counseling centers and reach individuals who otherwise would go without help. Summarizing, the emerging picture is that LLM chatbots are not a panacea or replacement for therapists, but when used wisely, they can complement human providers, extending support to times and settings where humans can't be present.

Limitations. Despite the promising results, our study – and the current state of AI chatbots – has important limitations. First, as a *pilot trial*, our sample size was small, and participants were not randomly assigned to conditions. This raises the risk of bias: for example, those who chose to use the ChatBot might have been more motivated or tech-comfortable, which could partially explain their improvements. Without a randomized control, we cannot definitively attribute symptom changes to the ChatBot itself. Non-specific factors like *expectancy effects* (enthusiasm for a novel technical tool) or the additional attention given to the ChatBot group might have contributed. Therefore, our outcome findings are preliminary and hypothesis-generating, not confirmatory. We have interpreted the efficacy signals cautiously, using them primarily to estimate effect sizes and refine the intervention for a future definitive trial, rather than to make clinical claims. Another limitation is the *short duration* of the intervention and the lack of follow-up measures to assess if gains were persistent over time. User engagement with digital tools often declines over time; it is therefore plausible that, in the absence of novelty effects and research oversight, students may discontinue using the chatbot after several weeks. Long-term follow-ups (6 to 12 months) are needed to evaluate whether LLM-based support yields lasting improvements. Our sample was also limited in scope – primarily subclinical young adults at one university. This affects *generalizability*. The ChatBot may

need further personalization: for instance, older adults or those with severe mental illness who might have different usability challenges or safety risks. We deliberately focused on a low-risk group; therefore, our positive findings may not generalize to higher-risk settings. Current evidence suggests that human intervention is indispensable for high-risk cases. Additionally, cultural factors could influence chatbot acceptability – our ChatBot communicated in Romanian and assumed a relatively informal, friendly tone which suited our sample, but different languages or cultural contexts might require nuanced adjustments (as seen in Kang & Hong’s cross-lingual design addressing Korean cultural needs) [53]. Technological limitations of the LLM itself are another concern. While Claude 3.7-Sonnet is currently an advanced LLM, it can still produce errors (“hallucinations”) or overly generic advice despite the layer of CBT consistent system prompts provided in its knowledge database. For example, some participants wished the AI could offer more personalized guidance, echoing a common critique that current models, despite being adaptive, may lack the depth of insight of a trained human therapist. These are inherent limits that could impact efficacy and alliance: a response that is perceived as canned or slightly off-target can lead to disengagement. In our study, such instances were rare, but as we scale up to more users, ensuring response quality and individualization will be important. Another limitation was that we did not measure the *therapeutic alliance* from the client’s perspective or functional improvement indicators (e.g., academic performance, sleep quality). Future studies should include dedicated alliance measures (to better compare AI-human alliance vs human-human) or ecological momentary assessments to capture how the ChatBot impacts users’ daily functioning. Finally, from an implementation standpoint, unresolved questions about liability and ethics remain outside our scope but require attention. For instance, if a chatbot gives harmful advice or fails to prevent a crisis, who is responsible – the developers, the clinicians deploying it, or the institution? Regulatory frameworks (FDA, HIPAA, etc.) for AI in healthcare are still evolving. Until clear guidelines are in place, institutions may be hesitant to fully integrate such tools. We navigated this by keeping the chatbot in a research setting with clinician oversight, but a real-world rollout would demand rigorous testing, training for providers, and perhaps limiting the chatbot’s autonomy in sensitive situations. In summary, our pilot underscores both the potential and the limitations of LLM-assisted therapy. The approach appears feasible and acceptable, and it shows promise in improving clinical outcomes for the right users. Yet, significant work remains to address the biases, safety concerns, and practical constraints before this can become a standard care component. We view these limitations not as deterrents, but as design challenges for the next research phase.

Final Considerations and Future Directions

This study provided an important initial look at how an LLM ChatBot can augment an evidence-based therapy (UP) protocol for young adults. In many ways, the findings offer *proof-of-concept* that such a hybrid intervention is workable: students used the ChatBot, liked it, and benefited from it. These insights now inform the planning of a definitive randomized controlled trial. A future RCT will be designed to rigorously evaluate the ChatBot's efficacy and safety on a larger scale, incorporating the lessons learned here. For example, we will aim to recruit a sufficiently powered sample and include appropriate control conditions (e.g., UP therapy with vs. without the ChatBot) to isolate the chatbot's added value. Follow-up assessments will also be implemented to see if improvements are maintained over time. Based on our pilot, the RCT will emphasize *continued exploratory objectives* such as optimizing user engagement strategies and examining potential moderators of the outcome. Our future trials will incorporate more robust safety monitoring (e.g., real-time keyword flags for risk content), a refined consent process addressing privacy concerns, and more specific training for users to maximize effective use of the ChatBot. We could also plan to formally measure mediators like emotional expression and homework compliance to test the mechanism suggested by Yokotani et al. (i.e., that the chatbot's encouragement of negative emotion expression leads to symptom relief) [55]. If an AI-augmented UP program proves effective, it could be scaled up as a public health intervention for youth mental health. University counseling centers, which frequently operate under considerable resource constraints, might deploy chatbots as a first-line resource for students with mild symptoms or as an after-hours support system. This could free human counselors to focus on higher-risk cases, thus optimizing resources. On a broader scale, accessible chatbots might help reduce the incidence of subclinical issues developing into full disorders by providing timely coping tools and psychoeducation. Our findings contribute to a cautiously optimistic view that *LLM chatbots, when thoughtfully integrated, have the potential to "dramatically increase access to support" and serve as a valuable complement to traditional services*. They will not replace therapists – but they might become trusted assistants. Ultimately, the goal is to harness technology to reach more people with quality mental health support. The early successes reported are heartening, but they must be balanced against the clear need for continued research and oversight. Our study was an early step, and like other pilots in this nascent field, it had a limited scope. Therefore, we can assert that integrating an LLM-based chatbot into UP therapy for subclinical students seems feasible, acceptable, and potentially effective in reducing distress. This aligns with other recent pilot and RCT findings that AI chatbots can be efficacious, scalable, and well-liked tools in mental health care. At the same time, significant challenges around safety, ethics, and long-term impact remain. By adhering to high standards (both scientific and ethical) and

following a staged research approach (i.e., pilot trials, RCTs, implementation studies), we can address these challenges. If successful, this line of work could contribute to a paradigm of technology-enhanced psychotherapy where human clinicians and AI systems work concomitantly to expand the reach and effectiveness of mental health services. We are mindful of the risks, but also motivated by the possibility that, with proper care, generative AI could become a valuable complement to traditional mental health services, helping to fill gaps in care and provide timely help to those in need. The present study offers cautious optimism that we are moving in that direction.

References:

1. Santomauro DF, Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, Abbafati C, Adolph C, Amlag JO, Aravkin AY, Bang-Jensen BL. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*. 2021;398(10312):1700-12. doi: 10.1016/s0140-6736(21)02143-7
2. World Health Organization. World mental health report: Transforming mental health for all. 2022. <https://www.who.int/publications/i/item/9789240049338> [accessed July 26, 2025].
3. Lu W, Bessaha M, Muñoz-Laboy M. Examination of young US adults' reasons for not seeking mental health care for depression, 2011-2019. *JAMA Network Open*. 2022;5(5):e2211393-. doi: 10.1001/jamanetworkopen.2022.11393
4. Cuijpers P, Karyotaki E, Eckshtain D, Ng MY, Corteselli KA, Noma H, Quero S, Weisz JR. Psychotherapy for depression across different age groups: a systematic review and meta-analysis. *JAMA psychiatry*. 2020;77(7):694-702. doi:10.1001/jamapsychiatry.2020.0164
5. National Institute for Health and Care Excellence (NICE). Depression in adults: treatment and management. 2022. <https://www.nice.org.uk/guidance/ng222> [accessed July 26, 2025].
6. Kazantzis N, Whittington C, Dattilio F. Meta-analysis of homework effects in cognitive and behavioral therapy: A replication and extension. *Clinical Psychology: Science and Practice*. 2010;17(2):144. doi: 10.1111/j.1468-2850.2010.01204.x
7. Fernandez E, Woldgabreal Y, Day A, Pham T, Gleich B, Aboujaoude E. Live psychotherapy by video versus in-person: A meta-analysis of efficacy and its relationship to types and targets of treatment. *Clinical Psychology & Psychotherapy*. 2021;28(6):1535-49. doi: 10.1002/cpp.2594
8. Cuijpers P, Miguel C, Harrer M, Plessen CY, Ciharova M, Ebert D, Karyotaki E. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: A comprehensive meta-analysis including 409 trials with

- 52,702 patients. *World Psychiatry*. 2023;22(1):105-15. doi: 10.1002/wps.21069
9. Glenn D, Golinelli D, Rose RD, Roy-Byrne P, Stein MB, Sullivan G, Bystritsky A, Sherbourne C, Craske MG. Who gets the most out of cognitive behavioral therapy for anxiety disorders? The role of treatment dose and patient engagement. *Journal of consulting and clinical psychology*. 2013;81(4):639. doi: 10.1037/a0033403
 10. Mausbach BT, Moore R, Roesch S, Cardenas V, Patterson TL. The relationship between homework compliance and therapy outcomes: An updated meta-analysis. *Cognitive therapy and research*. 2010;34(5):429-38. doi: 10.1007/s10608-010-9297-z
 11. Fanous M, Daniels J. What are the recovery and attrition outcomes for group CBT and individual CBT for generalised anxiety disorder in an IAPT service? An exploratory study. *The Cognitive Behaviour Therapist*. 2020;13:e15. doi: 10.1017/s1754470x20000045
 12. Demszky D, Yang D, Yeager DS, Bryan CJ, Clapper M, Chandhok S, Eichstaedt JC, Hecht C, Jamieson J, Johnson M, Jones M. Using large language models in psychology. *Nature Reviews Psychology*. 2023;2(11):688-701. doi: 10.1038/s44159-023-00241-5
 13. Lim SM, Shiau CW, Cheng LJ, Lau Y. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. *Behavior Therapy*. 2022;53(2):334-47. doi: 10.1016/j.beth.2021.09.007
 14. Li H, Zhang R, Lee YC, Kraut RE, Mohr DC. Systematic review and meta-analysis of AI-based conversational agents for promoting mental health and well-being. *NPJ Digital Medicine*. 2023;6(1):236. doi: 10.1038/s41746-023-00979-5
 15. Zhong W, Luo J, Zhang H. 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*. 2024;356:459-69. doi: 10.1016/j.jad.2024.04.057
 16. Hatch SG, Goodman ZT, Vowels L, Hatch HD, Brown AL, Guttman S, Le Y, Bailey B, Bailey RJ, Esplin CR, Harris SM. When ELIZA meets therapists: A Turing test for the heart and mind. *PLOS Mental Health*. 2025;2(2):e0000145. doi: 10.1371/journal.pmen.0000145
 17. McFadyen J, Habicht J, Dina LM, Harper R, Hauser TU, Rollwage M. AI-enabled conversational agent increases engagement with cognitive-behavioral therapy: A randomized controlled trial. *medRxiv*. Preprint posted online November 2, 2024. doi: 10.1101/2024.11.01.24316565
 18. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *The Canadian*

- Journal of Psychiatry. 2019;64(7):456-64. doi: 10.1177/0706743719828977
19. Radziwill NM, Benton MC. Evaluating quality of chatbots and intelligent conversational agents. arXiv. Preprint posted online April 15, 2017. doi: 10.48550/arXiv.1704.04579
 20. Kuhail MA, Alturki N, Thomas J, Alkhalifa AK, Alshardan A. Human-human vs human-AI therapy: An empirical study. *International Journal of Human-Computer Interaction*. 2025;41(11):6841-52. doi: 10.1080/10447318.2024.2385001
 21. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *Journal of medical Internet research*. 2020;22(7):e16021. doi: 10.2196/16021
 22. Khawaja Z, Bélisle-Pipon JC. Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Frontiers in Digital Health*. 2023;5:1278186. doi: 10.3389/fdgth.2023.1278186
 23. Ma Z, Mei Y, Su Z. Understanding the benefits and challenges of using large language model-based conversational agents for mental well-being support. *AMIA Annual Symposium Proceedings*; 2024 Jan 11. PMID: 38222348
 24. Babu A, Joseph AP. Digital wellness or digital dependency? a critical examination of mental health apps and their implications. *Frontiers in Psychiatry*. 2025;16:1581779. doi: 10.3389/fpsyt.2025.1581779
 25. Williams R, Hopkins S, Frampton C, Holt-Quick C, Merry SN, Stasiak K. 21-day stress detox: open trial of a universal well-being chatbot for young adults. *Social Sciences*. 2021 Oct 30;10(11):416. doi: 10.3390/socsci10110416
 26. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. Emma: An emotion-aware wellbeing chatbot. 8th international conference on affective computing and intelligent interaction (ACII); 2019 Sep 3; IEEE. doi: 10.1109/ACII.2019.8925455
 27. Sabour S, Zhang W, Xiao X, Zhang Y, Zheng Y, Wen J, Zhao J, Huang M. A chatbot for mental health support: exploring the impact of Emohaa on reducing mental distress in China. *Frontiers in digital health*. 2023;5:1133987. doi: 10.3389/fdgth.2023.1133987
 28. Melo A, Silva I, Lopes J. Chatgpt: A pilot study on a promising tool for mental health support in psychiatric inpatient care. *International Journal of Psychiatric Trainees*. 2024;2(2). doi: 10.55922/001c.92367
 29. Fitzpatrick KK, Darcy A, Vierhile M. Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial *JMIR Ment Health*. 2017;4(2):e19 doi:

[10.2196/mental.7785](https://doi.org/10.2196/mental.7785)

30. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial Intelligence–Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial. *JMIR Form Res* 2021;5(8):e20678. doi: [10.2196/20678](https://doi.org/10.2196/20678)
31. Suharwardy S, Ramachandran M, Leonard SA, Gunaseelan A, Lyell DJ, Darcy A, Robinson A, Judy A. Feasibility and impact of a mental health chatbot on postpartum mental health: a randomized controlled trial. *AJOG Global Reports*. 2023;3(3):100165. doi: 10.1016/j.xagr.2023.100165
32. MacNeill AL, Doucet S, Luke A. Effectiveness of a Mental Health Chatbot for People With Chronic Diseases: Randomized Controlled Trial. *JMIR Form Res* 2024;8:e50025. doi: [10.2196/50025](https://doi.org/10.2196/50025)
33. He Y, Yang L, Zhu X, Wu B, Zhang S, Qian C, Tian T. Mental health chatbot for young adults with depressive symptoms during the COVID-19 pandemic: single-blind, three-arm randomized controlled trial. *Journal of medical Internet research*. 2022;24(11):e40719. doi: 10.2196/40719
34. Habicht J, Dina LM, McFadyen J, Stylianou M, Harper R, Hauser TU, Rollwage M. Generative AI-enabled therapy support tool for improved clinical outcomes and patient engagement in group therapy: real-world observational study. *Journal of medical Internet research*. 2025;27:e60435. doi: 10.2196/60435
35. Jiang M, Zhao Q, Li J, Wang F, He T, Cheng X, Yang BX, Ho GW, Fu G. A generic review of integrating artificial intelligence in cognitive behavioral therapy. *arXiv*. Preprint posted online July 28, 2024. doi: 10.48550/arXiv.2407.19422
36. Wah JN. Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions. *Frontiers in Public Health*. 2025;13:1530799. doi: 10.3389/fpubh.2025.1530799
37. Han Q, Zhao C. Unleashing the potential of chatbots in mental health: bibliometric analysis. *Frontiers in Psychiatry*. 2025;16:1494355. doi: 10.3389/fpsy.2025.1494355
38. Barlow DH, Farchione TJ, Bullis JR, Gallagher MW, Murray-Latin H, Sauer-Zavala S, Bentley KH, Thompson-Hollands J, Conklin LR, Boswell JF, Ametaj A. The unified protocol for transdiagnostic treatment of emotional disorders compared with diagnosis-specific protocols for anxiety disorders: A randomized clinical trial. *JAMA psychiatry*. 2017;74(9):875-84. doi:10.1001/jamapsychiatry.2017.2164
39. Farchione TJ, Fairholme CP, Ellard KK, Boisseau CL, Thompson-Hollands J, Carl JR,

- Gallagher MW, Barlow DH. Unified protocol for transdiagnostic treatment of emotional disorders: a randomized controlled trial. *Behavior therapy*. 2012;43(3):666-78. doi: 10.1016/j.beth.2012.01.001
40. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj*. 2008;337. doi: 10.1136/bmj.a1655
41. Brooke J. SUS-A quick and dirty usability scale. *Usability evaluation in industry*. 1996;189(194):4-7. doi: 10.1201/9781498710411-35
42. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. *Intl. Journal of Human-Computer Interaction*. 2008;24(6):574-94. doi: 10.1080/10447310802205776
43. Connor KM, Davidson JR, Churchill LE, Sherwood A, Weisler RH, Foa E. Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. *The British Journal of Psychiatry*. 2000;176(4):379-86. doi: 10.1192/bjp.176.4.379
44. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of internal medicine*. 2006;166(10):1092-7. doi: 10.1001/archinte.166.10.1092
45. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*. 2001;16(9):606-13. doi: 10.1046/j.1525-1497.2001.016009606.x
46. Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. Internal construct validity of the Warwick-Edinburgh mental well-being scale (WEMWBS): a Rasch analysis using data from the Scottish health education population survey. *Health and quality of life outcomes*. 2009;7(1):15. doi: 10.1186/1477-7525-7-15
47. Ng Fat L, Scholes S, Boniface S, Mindell J, Stewart-Brown S. Evaluating and establishing national norms for mental wellbeing using the short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. *Quality of Life Research*. 2017;26(5):1129-44. doi: 10.1007/s11136-016-1454-8
48. Tulbure BT, Rusu A, Sava FA, Sălăgean N, Farchione TJ. A Web-Based Transdiagnostic Intervention for Affective and Mood Disorders: Randomized Controlled Trial. *JMIR Ment Health* 2018;5(2):e36. doi: [10.2196/mental.8901](https://doi.org/10.2196/mental.8901)
49. Isbăşoiu AB, Tulbure BT, Rusu A, Sava FA. Can we boost treatment adherence to an online Transdiagnostic intervention by adding self-enhancement strategies? Results from a

- randomized controlled non-inferiority trial. *Frontiers in psychology*. 2021;12:752249. doi: 10.3389/fpsyg.2021.752249
50. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative research in psychology*. 2006;3(2):77-101. doi: 10.1191/1478088706qp063oa
51. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical methods in medical research*. 2016;25(3):1057-73. doi: 10.1177/0962280215588241
52. Heinz MV, Mackin DM, Trudeau BM, Bhattacharya S, Wang Y, Banta HA, Jewett AD, Salzhauer AJ, Griffin TZ, Jacobson NC. Randomized trial of a generative AI chatbot for mental health treatment. *Nejm Ai*. 2025;2(4):AIoa2400802. doi: [10.1056/AIoa2400802](https://doi.org/10.1056/AIoa2400802)
53. Kang B, Hong M. Development and evaluation of a mental health chatbot using ChatGPT 4.0: Mixed methods user experience study with Korean users. *JMIR Medical Informatics*. 2025;13:e63538. doi: 10.2196/63538
54. Moore J, Grabb D, Agnew W, Klyman K, Chancellor S, Ong DC, Haber N. Expressing stigma and inappropriate responses prevents LLMs from safely replacing mental health providers. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 2025 Jun 23. doi: [10.1145/3715275.3732039](https://doi.org/10.1145/3715275.3732039)
55. Yokotani K, Ito M, Ihara N, Shigeeda Y. A unified protocol chatbot reduces anxiety by encouraging university students' negative emotional expressions: A randomized controlled trial. *Computers in Human Behavior Reports*. 2025:100770. doi: [10.1016/j.chbr.2025.100770](https://doi.org/10.1016/j.chbr.2025.100770)
56. Chen C, Lam KT, Yip KM, So HK, Lum TY, Wong IC, Yam JC, Chui CS, Ip P. Comparison of an AI Chatbot With a Nurse Hotline in Reducing Anxiety and Depression Levels in the General Population: Pilot Randomized Controlled Trial. *JMIR human factors*. 2025;12:e65785. doi: 10.2196/65785
57. Alanzi TM, Alharthi A, Alrumman S, Abanmi S, Jumah A, Alansari H, Alharthi T, Alibrahim A, Algethami A, Aburass M, Alshahrani AM. ChatGPT as a psychotherapist for anxiety disorders: An empirical study with anxiety patients. *Nutrition and Health*. 2024:02601060241281906. doi: [10.1177/02601060241281906](https://doi.org/10.1177/02601060241281906)
58. Coombs NC, Meriwether WE, Caringi J, Newcomer SR. Barriers to healthcare access among US adults with mental health challenges: A population-based study. *SSM-population health*. 2021;15:100847. doi: 10.1016/j.ssmph.2021.100847.
59. Akdogan O, Uyar GC, Yesilbas E, Baskurt K, Malkoc NA, Ozdemir N, Yazici O, Oksuzoglu

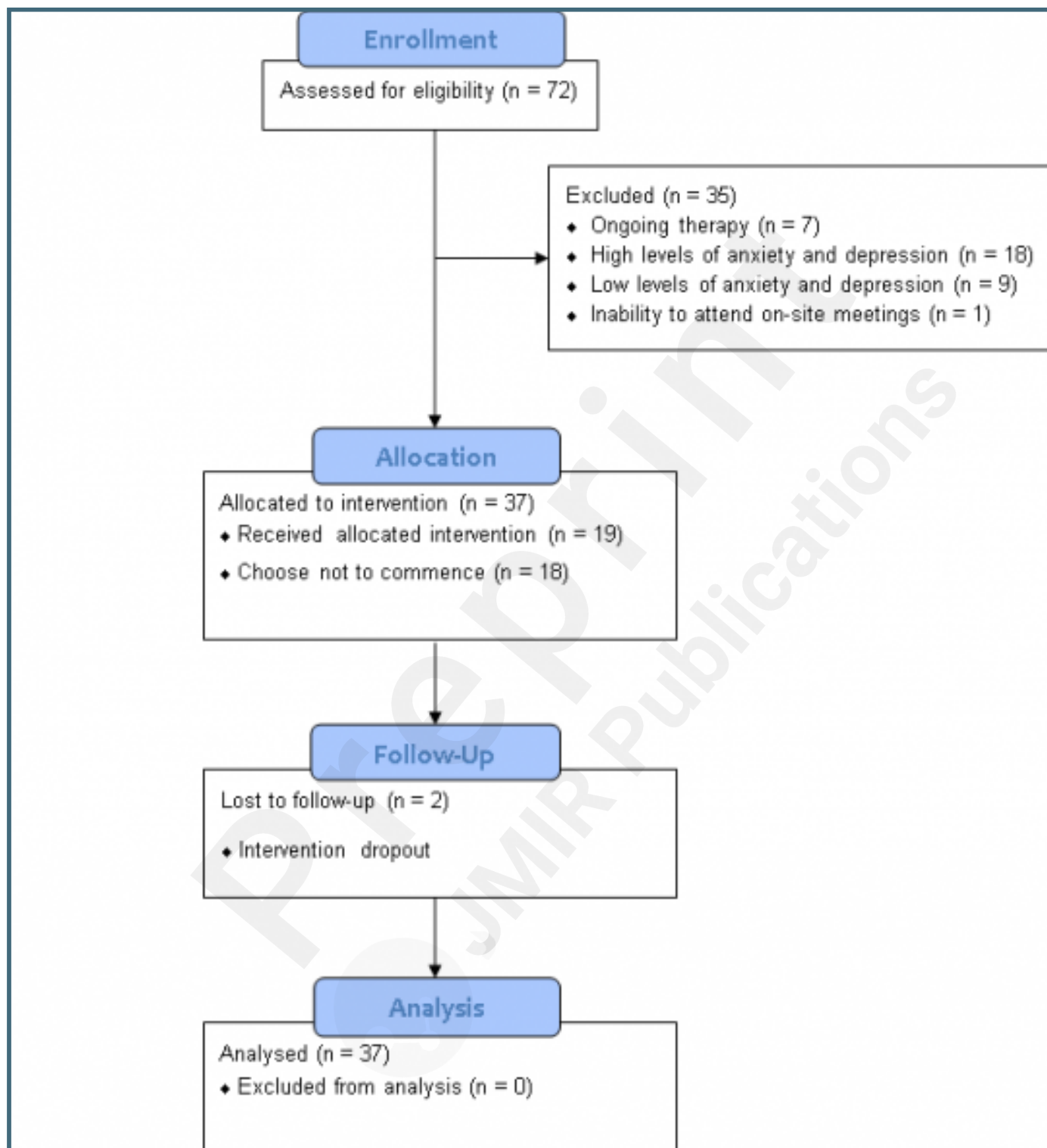
B, Uner A, Ozet A, Sutcuoglu O. Effect of a ChatGPT-based digital counseling intervention on anxiety and depression in patients with cancer: A prospective, randomized trial. *European Journal of Cancer*. 2025;221:115408. doi: 1016/j.ejca.2025.115408



Supplementary Files

Figures

Modified CONSORT flow diagram for participants' recruitment and progress throughout the program.



CONSORT (or other) checklists

Document containing the CONSORT checklist.

URL: <http://asset.jmir.pub/assets/d35e303d1933c2fd96d178867e80ab20.pdf>

