

Original Paper

# Assessing the Feasibility, Usability, Acceptability, and Efficacy of an AI Chatbot for Sleep Promotion: Quasi-Experimental Study

Xiaoyue Liu<sup>1</sup>, PhD; Jingchen Liu<sup>2</sup>, PhD

<sup>1</sup>New York University, New York, NY, United States

<sup>2</sup>Columbia University, New York, NY, United States

**Corresponding Author:**

Xiaoyue Liu, PhD  
New York University  
433 1st Ave  
New York, NY, 10010  
United States  
Phone: 1 2129925994  
Email: [xl5272@nyu.edu](mailto:xl5272@nyu.edu)

## Abstract

**Background:** Poor sleep is a concerning public health problem in the United States. Previous sleep interventions often face barriers such as high costs, limited accessibility, and low user engagement. Recent advancements in artificial intelligence (AI) technologies offer a novel approach to overcoming these limitations. In response, our team developed a prototype AI sleep chatbot powered by a large language model to deliver personalized, accessible sleep support.

**Objective:** This study aimed to examine the feasibility, usability, acceptability, and preliminary efficacy of the AI chatbot for sleep promotion.

**Methods:** We conducted a quasi-experimental, single-group study with adults in the United States aged 18 to 75 years who self-reported poor sleep. The chatbot was integrated into a commercially available messaging app. Participants were asked to engage with a virtual sleep therapist via texting over 2 weeks. The chatbot provided ongoing, individualized sleep guidance and adapted recommendations based on participants' prior conversations. Feasibility, usability, and acceptability were descriptively summarized. Sleep was assessed using questionnaires before and after the intervention.

**Results:** Of the 107 adults who enrolled in the study, 88 (82.2%) completed chatbot registration. Among these 88 participants, 65 (73.9%) initiated interactions, and 44 (50%) completed the 2-week intervention. The final analysis included 42 adults (mean age 36, SD 11 years; n=12, 28.6% male). On average, participants engaged with the chatbot for 58 (SD 42) minutes, with each chat session lasting approximately 9 (SD 6) minutes. Most reported favorable experiences with the chatbot. The average usability score was 85.2 (SD 10.7) out of 100, which was well above the benchmark of 68. The chatbot was rated as highly acceptable, with a satisfaction score of 27.3 (SD 4.1) out of 32. All participants perceived the chatbot as effective, with ratings ranging from "slightly effective" to "extremely effective." The preliminary evidence showed improved sleep outcomes after chatbot use: total sleep time increased by 1.4 hours ( $P<.001$ ); sleep onset latency decreased by 30.9 minutes ( $P<.001$ ); sleep efficiency increased by 7.8% ( $P=.007$ ); and scores improved for perceived sleep quality (mean difference [MD] -5.4;  $P<.001$ ), insomnia severity (MD -7.9;  $P<.001$ ), daytime sleepiness (MD -4.7;  $P<.001$ ), and sleep hygiene skills (MD -13.2;  $P<.001$ ). No significant change was observed in sleep environment (MD -1.1;  $P=.16$ ).

**Conclusions:** Our AI chatbot demonstrated satisfactory feasibility, usability, and acceptability. Improvements were observed following chatbot use, although causality cannot be established. These findings highlight the potential of integrating state-of-the-art large language models into behavioral interventions for sleep promotion. Future research should include objective sleep measurements and conduct randomized controlled trials to validate the study findings. If confirmed, this AI chatbot could be scaled to support sleep health on a broader level.

(JMIR Form Res 2026;10:e84023) doi: [10.2196/84023](https://doi.org/10.2196/84023)

**KEYWORDS**

artificial intelligence; AI; large language models; chatbot; feasibility; usability; acceptability; sleep

## Introduction

Poor sleep is a major public health concern in the United States. Major guidelines recommend that adults sleep at least 7 hours per night [1,2]. Over 25% of adults in the United States do not meet this recommendation, and 14.5% had difficulty falling asleep on most days or every day over the previous month [3]. Deviations from optimal sleep duration and quality play a crucial role in the development and progression of various health conditions. For example, adults with chronic experience of poor sleep have a higher risk of developing depression later in life [4]. Similarly, extensive epidemiological evidence links insufficient sleep to an increased risk of cardiovascular disease [5-7]. Adults with poor sleep quality have a 1.38-fold higher mortality rate in comparison to those with good sleep quality [8]. The ramifications extend beyond individual health. Poor sleep directly links to reduced productivity, lower work performance, and increased health care use [9]. On average, poor sleep is associated with US \$3400 to US \$5200 in additional health care expenditures per person [10]. At the population level, insufficient sleep causes the United States to lose approximately 1.23 million workdays annually, with projected economic costs of up to US \$456 billion by 2030 [11]. Despite these serious consequences, health care providers rarely target sleep as an approach to improve health outcomes [12]. Therefore, implementing an effective intervention to alleviate the public health burden resulting from poor sleep is a must.

Existing behavioral interventions such as cognitive behavioral therapy for insomnia (CBT-I) have proven effective in improving sleep outcomes. However, their widespread use is primarily constrained by high costs, limited accessibility, and a lack of personalization. These barriers hinder the broad adoption of sleep interventions. Consequently, digital tools have emerged as a solution for enhancing sleep health. There has been growing interest in incorporating artificial intelligence (AI) into treatments. AI-based technologies, particularly conversational tools, have shown growing evidence of health behavior change. A recent systematic review found that chatbots can promote healthy lifestyles, support smoking cessation, improve medication adherence, and reduce substance use [13]. By simulating humanlike conversations, chatbots deliver on-demand, tailored support that is typically not achievable with traditional in-person or telehealth care. Although limited, emerging research has suggested the potential benefit of using chatbots in improving sleep [14]. Nurses who used a chatbot that offered care support and coping assistance reported better sleep quality after 6 weeks in comparison to the control group [15]. Another study used a chatbot to gather conversations about sleep from parents of preterm and term infants to better understand their personal experiences [16]. Yet, most chatbot interventions still rely on decision tree algorithms, which often struggle to interpret conversational nuances and context. Recent breakthroughs in generative AI offer novel alternatives to conventional behavioral interventions. Generative AI systems learn patterns from large datasets to produce new, realistic, and coherent content that resembles human communication. A prominent example is large language models (LLMs), an advanced form of generative AI that uses deep learning to

comprehend and generate natural-sounding responses. Currently, there is scarce or no evidence on integrating state-of-the-art LLMs in sleep interventions. In response, we developed a prototype AI chatbot using LLMs that provides responses specifically to promote sleep. This study sought to (1) evaluate the feasibility, usability, and acceptability of the AI chatbot; and (2) test the preliminary efficacy of the AI chatbot on improving sleep outcomes.

## Methods

### Study Design and Sample

We used a quasi-experimental 1-group design. Using a convenience sampling approach, participants who self-reported poor sleep in the previous month were recruited nationwide in the United States. Our inclusion criteria were (1) age of 18 to 75 years, (2) a raw score of 23 or higher on the Patient-Reported Outcomes Measurement Information System Sleep Disturbance [17], (3) ownership of a smartphone and ability to access the internet, (4) willingness and ability to participate through texting on a smartphone, and (5) ability to read and understand English. Individuals were excluded if they were pregnant, lactating, or concurrently participating in other research studies focused on lifestyle modifications.

### Generative AI Technology

Figure 1 illustrates the automated workflow of the generative AI technology. The main body of the infrastructure comprises a central app designed by our study team, which communicates with a messaging app and LLMs via an application programming interface.

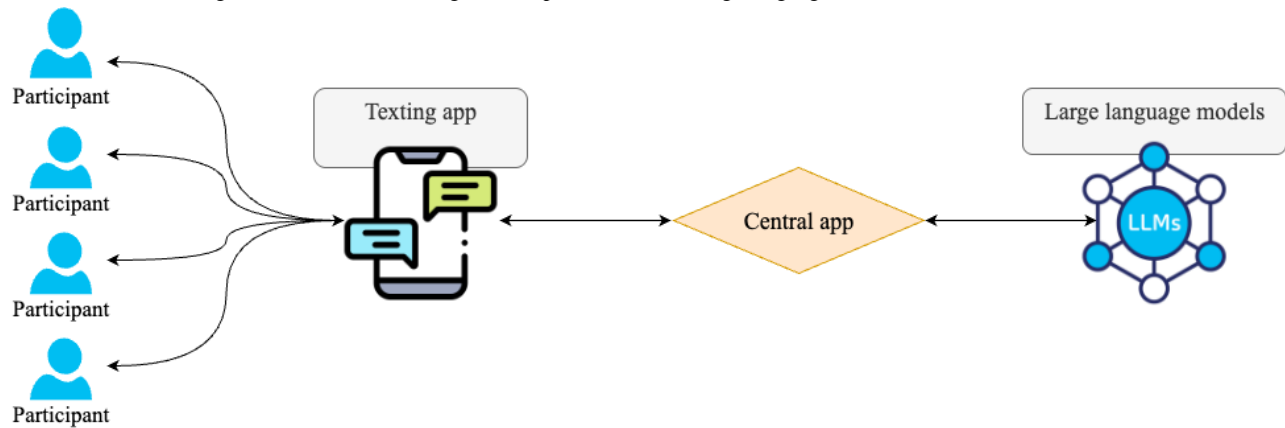
We created an AI chatbot on the messaging app platform that allowed participants to exchange texts as if they were interacting with a real sleep therapist. The chat operates on a commercially available messaging platform, which has a wide user base in the United States and is available to both iOS and Android smartphones. When a participant sends a message to the chatbot, the message is pushed to the central app. The central app is entirely controlled by our study team. It processes incoming messages based on the participant's chat history and then sends a query to the LLMs. We used Gemini-1.5-pro-001 (Google) and the CBT-I manual as the context for each prompt, with guidance focused on the following treatment components: stimulus control (strengthen the bed as a cue for sleep), sleep hygiene practices (promote healthy sleep habits), relaxation training (improve techniques that relax the mind and body), and cognitive restructuring (change negative thoughts and beliefs about sleep). For safety concerns, the prompts were designed to exclude sleep restriction therapy, a key CBT-I element that involves reducing time in bed. The responses from the LLMs then return to the central app, which subsequently forwards them to the participant through the chatbot on the messaging app.

To ensure participant safety, we implemented several safeguards and configured the LLMs with a reasonably high safety setting. Details of the model configuration and safety parameters are provided in [Multimedia Appendix 1](#). All conversational inputs and outputs were screened for hate speech, harassment, sexual

content, self-harm, or medical misinformation. Before deployment, we prompted the LLMs to avoid providing medical diagnoses or medication advice and stress tested them with adversarial prompts. Participants were instructed not to share personal identifiers when interacting with the chatbot. Any inadvertent disclosures were immediately removed from the chat history by the study team. Communications with the LLMs

and with the third-party messaging app were encrypted in transit using HTTPS. Our team monitored the conversational dialogues daily throughout the intervention phase to ensure that the advice delivered to participants was appropriate and safe, and to promptly identify any conversations necessitating immediate action (eg, indication of self-harm).

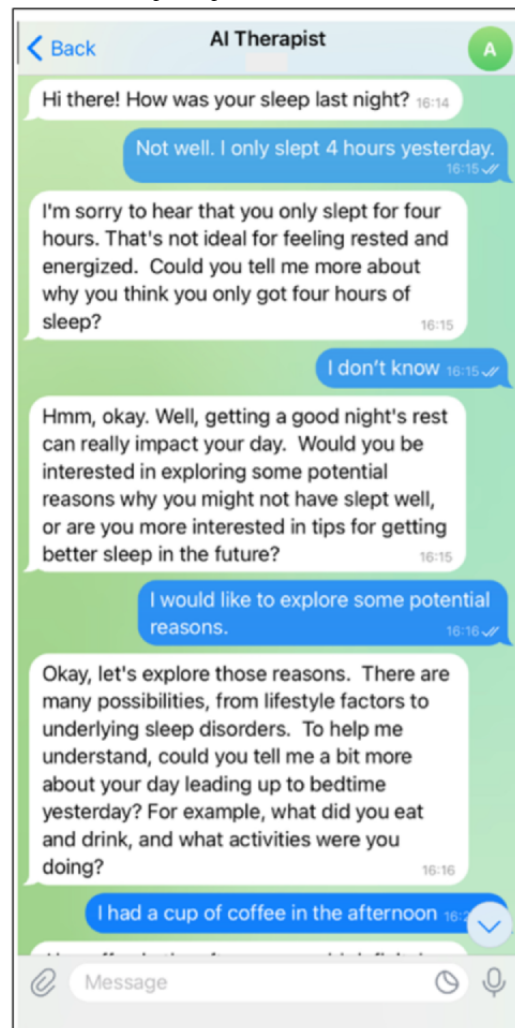
**Figure 1.** Workflow of the generative artificial intelligence sleep chatbot. LLM: large language model.



### Intervention Description

The intervention lasted 2 weeks and was delivered through the chatbot designed to function as a virtual sleep therapist. It created individualized treatment plans tailored to each participant's needs and feedback. By applying motivational interviewing techniques, the chatbot explored underlying causes of sleep disturbances and identified appropriate solutions aligned

with key components of CBT-I. If certain issues were unresolved, the chatbot conducted further analysis using LLMs to examine additional strategies. The chatbot also retained memory of previous conversations, allowing it to follow up with participants, reinforce recommendations, and detect any challenges in adherence. An example of conversations between the AI chatbot and a participant is illustrated in [Figure 2](#).

**Figure 2.** Example of conversations between the chatbot and participant.

## Ethical Considerations

This study was approved by the New York University Institutional Review Board (FY2024-9037). Informed consent was obtained from all participants before any procedures. Participation was voluntary. Individuals could withdraw at any time or skip any question without penalty. Each participant was assigned a unique study ID. Data were securely stored on university-managed, access-controlled devices and systems with access limited to study personnel. The university devices and systems encrypt the data at rest. We did not have a data processing agreement with the third-party messaging app. However, participants were informed during the consent procedures that the study required the use of a third-party messaging platform. All personal identifiers were removed before analysis. Deidentified analytic files will be retained for 5 years after publication and then permanently deleted.

## Procedures

After providing informed consent, participants registered for the chatbot on a third-party messaging app. They then completed a prestudy survey that inquired about their sociodemographic information, clinical background, and sleep characteristics. The intervention began when participants initiated conversations with the chatbot via texting. To protect confidentiality and

privacy, the chatbot only responded to those who initiated contact. Participants were instructed to text with the chatbot at least every 3 days for 2 weeks, with each chat session lasting approximately 3 to 5 minutes. We classified individuals as dropouts if they interacted with the chatbot fewer than 4 days in total or had an average daily duration of engagement of less than 1 minute during the intervention period. In this study, participants could chat as frequently or as long as they preferred. After the initial conversations, the chatbot followed up with participants if no interactions were detected within 72 hours. In subsequent sessions, the chatbot checked in with participants about their sleep concerns, reinforced recommended sleep promotion strategies, and suggested new ones when necessary. Finally, participants completed a poststudy survey upon concluding the intervention.

## Measurements

### Participant Flow

Feasibility was assessed based on the following metrics: (1) the proportion of individuals who registered for the chatbot; (2) the proportion of registered participants who initiated engagement with the chatbot; (3) the proportion of active participants who completed the 2-week intervention; and (4) chatbot engagement, including the total number of days interacting with the chatbot, total duration of chat sessions, and average duration per session.



## Usability

Usability was measured via the Chatbot Usability Questionnaire (CUQ) [18]. This instrument consists of 16 items scored on a 5-point Likert-type scale ranging from 1 (“strongly disagree”) to 5 (“strongly agree”). It measures personality, onboarding, user experience, and error handling of a chatbot. Odd-numbered items assess positive aspects, whereas even-numbered items assess negative aspects. Responses to all items are summed and then converted to a 100-point scale to be comparable with the System Usability Scale [19]. A total score of 68 is typically used as the threshold for acceptable usability [20].

## Acceptability

Acceptability was measured using the adapted Client Satisfaction Questionnaire (CSQ) [21] and 6 questions developed by the study team. The original CSQ is an instrument to assess a client’s overall level of satisfaction [21]. We revised the wording to make the questions fit with the study context. Each item is rated on a 4-point Likert scale from 1 (“poor”) to 4 (“excellent”). The total score ranges from 8 to 32, with a higher number indicating greater satisfaction. We also developed the following questions to evaluate the participants’ experiences: (1) “What changes have you noticed about your sleep after engaging with the chatbot?”; (2) “How effective were the chatbot’s tips for improving your sleep?” (rated from 1=“Not effective” to 5=“Extremely effective”); (3) “Did the chatbot offer personalized advice based on your input?” (rated from 1=“Never” to 5=“Always personalized”); (4) “How well did the chatbot adapt to your sleep concerns?” (rated from 1=“Not at all” to 5=“Very well”); (5) “How well did the chatbot remember your preferences in follow-up conversations?” (rated from 1=“Not at all” to 5=“Very well”); and (6) “Did you feel supported by the chatbot during your interactions?” (rated from 1=“Not at all” to 5=“Very well”). These self-developed items were validated scales, and they were intended only for formative evaluation.

## Preliminary Efficacy

The preliminary efficacy of the chatbot was examined using several metrics.

Habitual sleep patterns and perceived sleep quality were measured using the Pittsburgh Sleep Quality Index [22]. This 19-item instrument evaluates sleep quality and disturbances over the previous month. The items cover 7 components: subjective sleep quality, sleep onset, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. Each component score ranges from 0 to 3. These components are then summed to generate a global score ranging from 0 to 21, where a cutoff score of 5 or higher indicates poor sleep quality. Additionally, we used items from the Pittsburgh Sleep Quality Index to assess 3 habitual sleep patterns, which included sleep onset, sleep duration, and sleep efficiency. Sleep onset was measured using the question “How long has it usually taken you to fall asleep each night?” Sleep duration was measured using the question “How many hours of actual sleep did you get at night?” Sleep efficiency was calculated by dividing sleep duration by time spent in bed and multiplying the result by 100.

Insomnia severity was measured using the Insomnia Severity Index [23]. The Insomnia Severity Index assesses the severity and impact of insomnia over the previous 2 weeks. It comprises 7 items, with the score of each item evaluated on a 5-point Likert-type scale. The total score ranges from 0 to 28, and a higher score indicates more severe insomnia symptoms. A total score over 14 indicates clinical insomnia.

Daytime sleepiness was measured using the Epworth Sleepiness Scale. The Epworth Sleepiness Scale is an 8-item, 4-point Likert scale that measures the likelihood of dozing off or falling asleep in 8 common life situations [24]. Participants who had not experienced some of the situations were asked to estimate how each might affect them. A total score over 10 indicates excessive daytime sleepiness.

Sleep hygiene was measured using the Sleep Hygiene Index [25]. The Sleep Hygiene Index is a 13-item, 5-point Likert scale to evaluate the frequency of participants engaging in specific sleep behaviors. The sum of each item score yields a global score ranging from 0 to 52, with a higher score representing more maladaptive behaviors that compromise sleep hygiene.

Sleep environment was measured using the Assessment of Sleep Environment [26]. This instrument includes 13 items that quantify the impact of light and dark, noise, smell, humidity, comfort of the sleeping surface and pillows, and safety on an individual’s sleep. Each item is rated on a scale from 0 (“strongly disagree”) to 3 (“strongly agree”) and can be summed to obtain a total score. A higher total score suggests living in a poorer sleep environment.

## Statistical Analysis

Sociodemographic and clinical characteristics were summarized as mean and SD for continuous variables or as count and percentage for categorical variables. Conversational dialogue data were analyzed descriptively to assess the level of engagement. Usability and acceptability measures were used to visualize bar plots. Changes in each sleep measure before and after the intervention were assessed for normality using the Shapiro-Wilk test and visualized using histograms and *Q-Q* plots. Outliers were identified and removed. Depending on whether normality assumptions were met, pretest-posttest differences in each sleep variable were evaluated using either 2-tailed paired *t* tests or Wilcoxon signed rank tests. We also computed the Cohen *d* to estimate the effect size associated with each pretest-posttest change. Additionally, we used scatterplots to visualize individual changes in sleep variables, with posttest values on the x-axis and pretest values on the y-axis. We also calculated the proportion of participants who showed improvements in each sleep measure after the intervention. All statistical analyses were performed using R (version 4.5.0; R Foundation for Statistical Computing) [27]. The significance threshold was set at  $\alpha=.05$ .

## Results

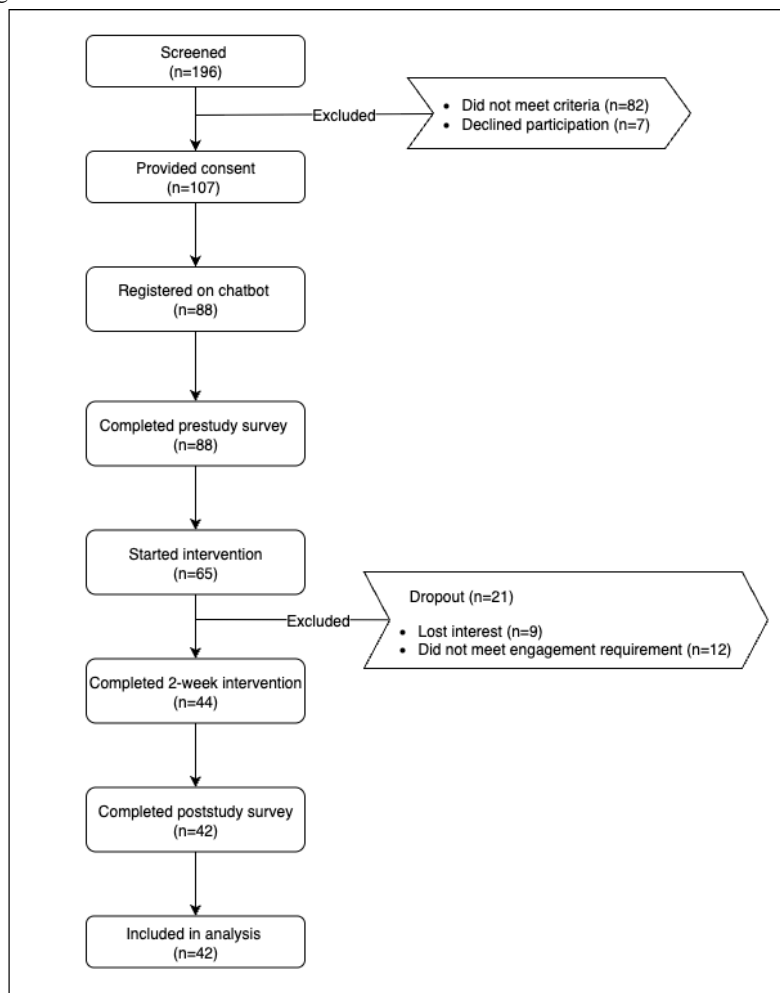
### Feasibility

Figure 3 presents the participant flow diagram for this study. A total of 196 individuals were screened for eligibility. Of those 196 individuals, 107 (54.6%) adults met the study criteria and

provided informed consent. Of these 107 adults, 88 (82.2%) completed chatbot registration and the prestudy survey. A total of 60.7% (65/107) of the participants initiated conversations with the chatbot. In total, 41.1% (44/107) of the participants completed the 2-week intervention. Of these 44 participants, 42 (95.5%) completed the poststudy survey and were included

in the final analysis. The completion rates for all outcome variables were 100% without missing data. In [Multimedia Appendix 2](#), we compare baseline characteristics between participants who completed the intervention and those who did not start it or did not complete it. No statistically significant differences were found between the 2 groups.

**Figure 3.** Participant flow diagram.



### Sample Characteristics

[Table 1](#) summarizes the key sociodemographic and clinical characteristics of the 42 study participants. Their mean age was 36 (SD 11) years, with a range from 22 to 59 years. A total of 28.6% (12/42) of the participants were male, and 61.9% (26/42)

were White individuals. In total, 28.6% (12/42) reported taking sleep medications at night. A total of 81.0% (34/42) of the participants slept less than 7 hours per night. Nearly all (41/42, 97.6%) reported having poor sleep quality. In total, 73.8% (31/42) met the criteria for clinical insomnia, and 47.6% (20/42) experienced excessive daytime sleepiness.

**Table 1.** Characteristics of the study sample (n=42).

Characteristic	Values
<b>Sociodemographic and clinical characteristics</b>	
Age (y), mean (SD)	36 (11)
<b>Sex, n (%)</b>	
Male	12 (28.6)
Female	30 (71.4)
<b>Race, n (%)</b>	
Non-White	16 (38.1)
White	26 (61.9)
<b>Educational level, n (%)</b>	
Lower than a bachelor's degree	15 (35.7)
Bachelor's degree or higher	27 (64.3)
Employed, n (%)	24 (57.1)
Insured, n (%)	39 (92.9)
<b>Income (US \$), n (%)</b>	
≤50,000	18 (42.9)
50,001-99,999	14 (33.3)
≥100,000	10 (23.8)
<b>BMI (kg/m<sup>2</sup>), mean (SD)</b>	29.7 (8.2)
Normal, n (%)	11 (26.2)
Overweight, n (%)	15 (35.7)
Obese, n (%)	15 (35.7)
<b>Sleep characteristics, n (%)</b>	
Short sleep (<7 hours per night)	34 (81.0)
Poor sleep quality (Pittsburgh Sleep Quality Index>5)	41 (97.6)
Clinical insomnia (Insomnia Severity Index>14)	31 (73.8)
Excessive daytime sleepiness (Epworth Sleepiness Scale>10)	20 (47.6)
History of taking sleep medications	12 (28.6)

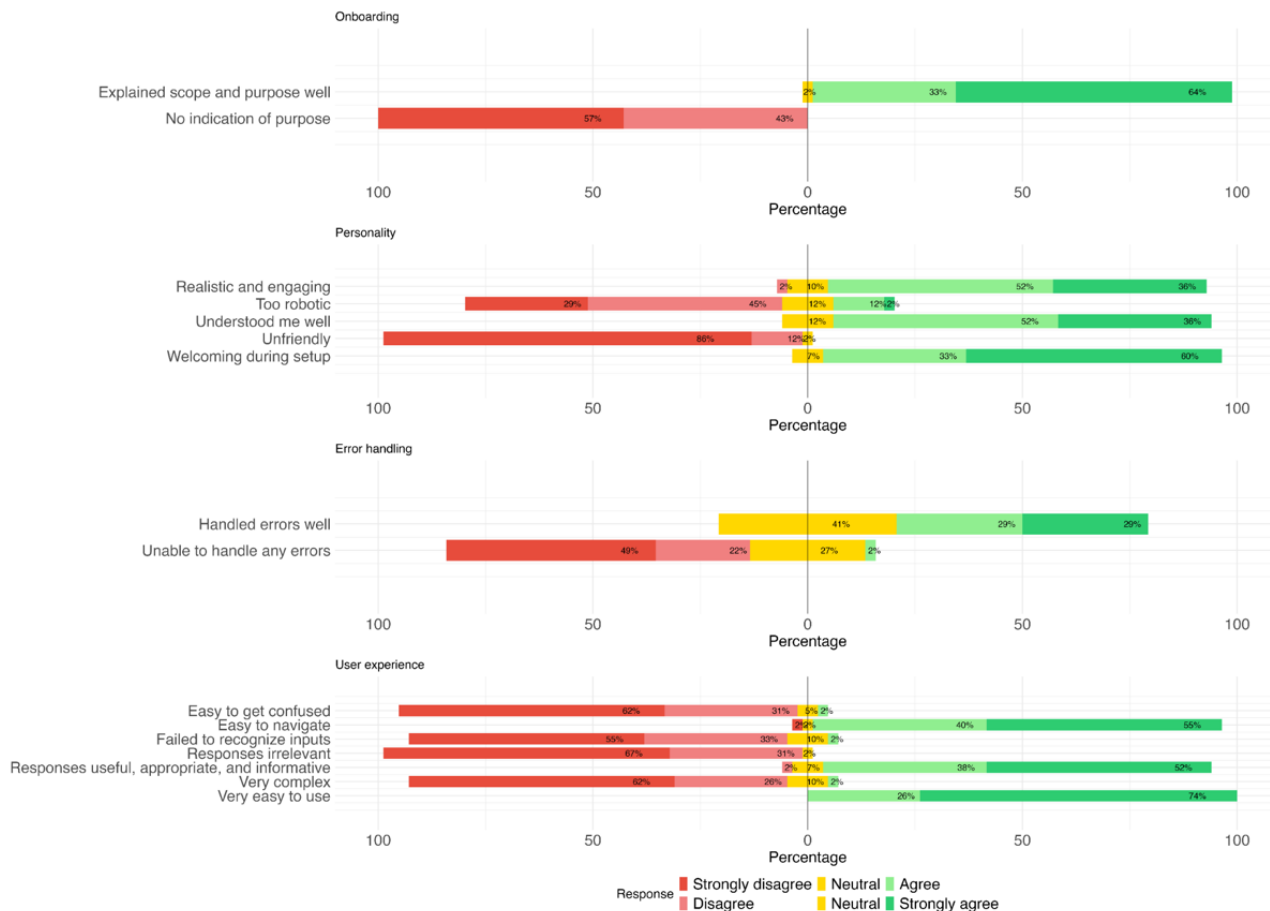
## Engagement With the Chatbot

Participants engaged with the chatbot for an average of 7 days over the 2-week intervention (median 7, IQR 5-8, range 4-13 days). On average, they spent 58 minutes interacting with the chatbot (median 49, IQR 27-69, range 9-201 minutes), with each chat session lasting approximately 9 minutes per day (median 7, IQR 4-11, range 1-36 minutes). Correlations between daily chatbot use and each outcome variable are presented in [Multimedia Appendix 3](#). In brief, satisfaction with the chatbot was the only variable that showed a strong correlation with daily use time ( $r=0.39$ ;  $P=.01$ ).

## Usability

The average CUQ score was 85.2 (SD 10.7), indicating a high level of usability compared to the benchmark score of 68.

Usability was further assessed across 4 CUQ domains ([Figure 4](#)). The onboarding experience received positive feedback, with 98% (41/42) of participants agreeing that the chatbot clearly explained its scope and purpose, and no one felt that the chatbot failed to indicate its purpose. Regarding personality, 88% (37/42) to 98% (41/42) agreed that the chatbot was realistic, welcoming, able to understand them, and friendly, although 14% (6/42) felt that it was somewhat robotic. Participants also reported favorable overall experiences, with 81% (34/42) to 100% (42/42) indicating that the chatbot was easy to navigate and use and that it provided useful and appropriate responses. Only 2% (1/42) found the chatbot complex. In contrast, responses regarding error handling were mixed. There were 57% (24/42) of participants indicating that the chatbot handled mistakes well, 40% (17/42) remained neutral, and 2% (1/42) believed that it was unable to manage errors effectively.

**Figure 4.** Survey responses to the Chatbot Usability Questionnaire.

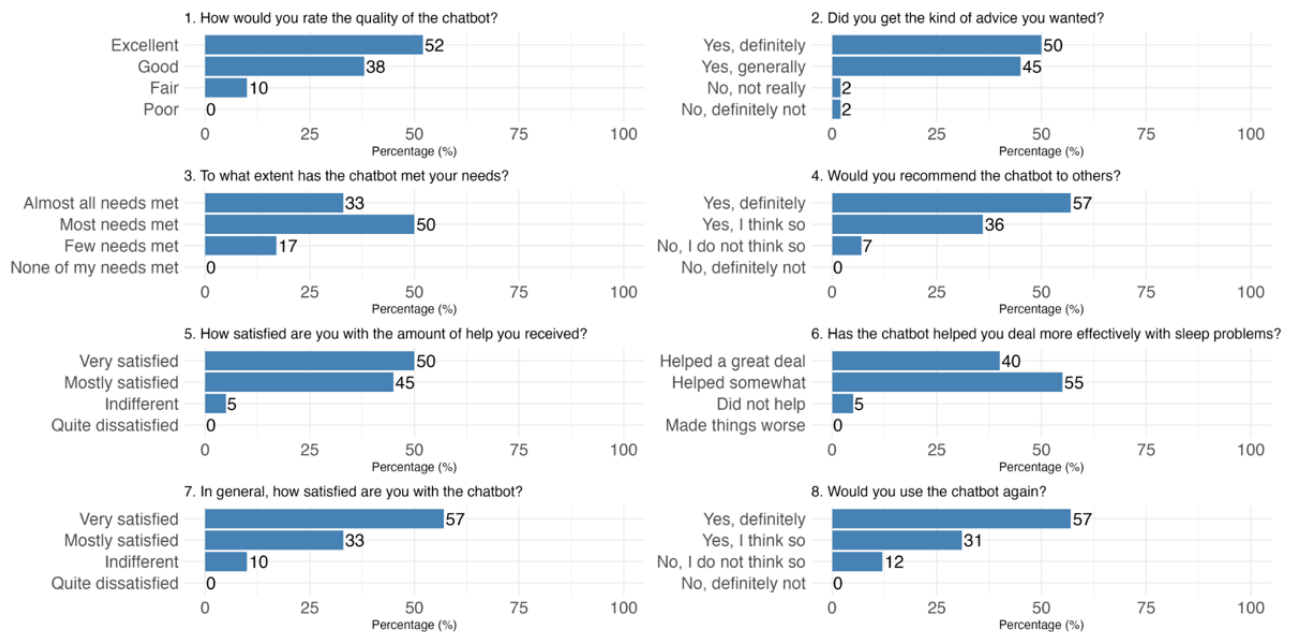
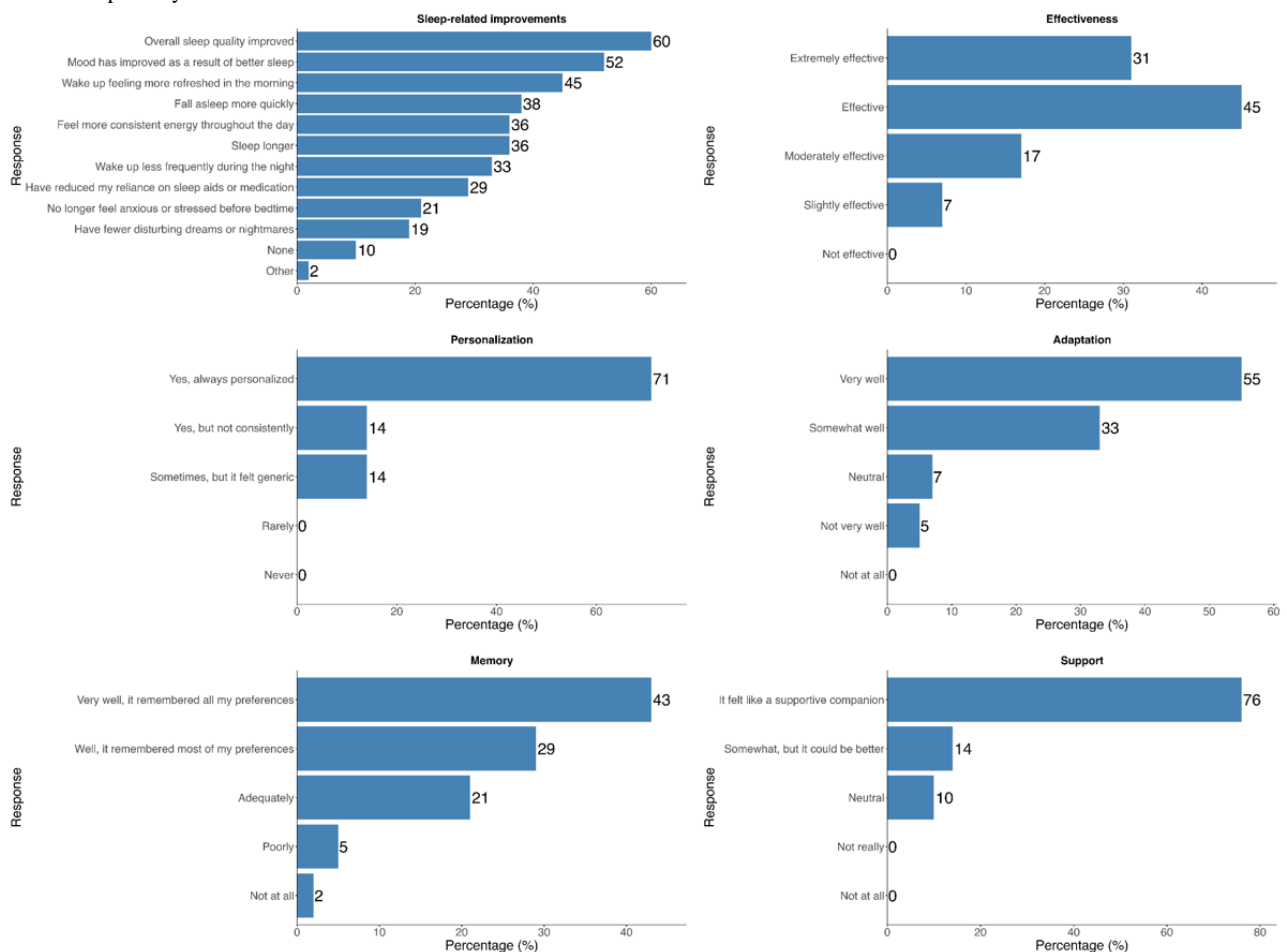
## Acceptability

The participants found the chatbot highly acceptable, providing an average adapted CSQ score of 27.3 (SD 4.1). Most (38/42, 90%) expressed overall satisfaction with the chatbot. The chatbot's educational content and advice were well received, with 95% (40/42) agreeing that it helped them manage sleep problems more effectively. Most participants (35/42, 83%) found that their needs were met: 50% (21/42) reported "most needs met," and 33% (14/42) reported "almost all needs met." In total, 88% (37/42) indicated that they would use the chatbot again, and 93% (39/42) would recommend it to others (Figure 5).

Participants expressed improvements in several sleep-related areas. The most frequently reported improvement was overall sleep quality (25/42, 60%), followed by mood enhancement (22/42, 52%) and feeling more refreshed in the morning (19/42,

45%). Over one-third reported falling asleep more quickly (16/42, 38%), experiencing more consistent energy throughout the day (15/42, 36%), sleeping longer (15/42, 36%), and waking less frequently during the night (14/42, 33%). All participants considered the chatbot effective, with ratings ranging from "slightly effective" to "extremely effective." A total of 71% (30/42) reported that the chatbot always personalized its advice, 14% (6/42) felt that personalization was inconsistent, and another 14% (6/42) perceived the responses as generic. Most participants (37/42, 88%) reported that the chatbot adapted its responses to their needs somewhat to very well; however, 5% (2/42) indicated that it did so not very well. A total of 93% (39/42) said that the chatbot was able to remember previous conversations, although 7% (3/42) reported poor or no recall. While most participants viewed the chatbot as a supportive companion, 24% (10/42) felt that its level of support could be better (Figure 6).



**Figure 5.** Survey responses to the Client Satisfaction Questionnaire.**Figure 6.** Acceptability of the chatbot.

### Preliminary Efficacy

The sample reported an average total sleep time of 5.3 hours, a sleep onset of 59.5 minutes, and a sleep efficiency of 76.2% at baseline. Following the intervention, changes were observed in

all 3 habitual sleep metrics: total sleep time increased by 1.4 hours ( $P < .001$ ), sleep onset decreased by 30.9 minutes ( $P < .001$ ), and sleep efficiency increased by 7.8% ( $P = .007$ ). We also found reductions in the scores of sleep quality (mean difference [MD] 5.4;  $P < .001$ ), insomnia severity (MD -7.9;  $P < .001$ ), daytime

sleepiness (MD  $-4.7$ ;  $P<.001$ ), and sleep hygiene practices (MD  $-13.2$ ;  $P<.001$ ). There was no statistically significant difference in sleep environment scores (MD  $-1.1$ ;  $P=.16$ ) before and after the intervention. The proportion of participants who showed

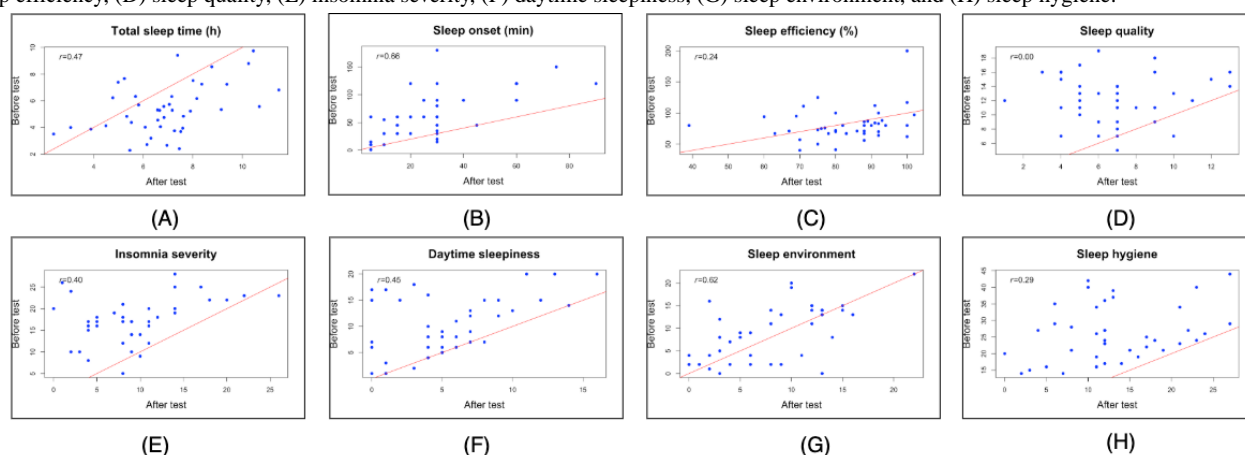
improvement in sleep measurements ranged from 60% (25/42; sleep environment) to 100% (42/42; sleep hygiene practices). The pre- and postintervention changes in sleep measurements are shown in Table 2 and Figure 7.

**Table 2.** Differences in sleep measurements before and after the intervention (n=42).

Sleep variables	Pretest time point, mean (SD)	Posttest time point, mean (SD)	Pretest-posttest differences			Participants improved, n (%)
			Mean difference (SD; 95% CI)	<i>P</i> value	Cohen <i>d</i>	
Nocturnal sleep metrics						
Total sleep time (h)	5.3 (1.5)	6.7 (1.3)	1.4 (1.4; 1.0 to 2.0)	<.001	0.99	30 (71.4)
Sleep onset <sup>a</sup> (min)	59.5 (38.3)	28.7 (20.0)	−30.9 (27.5; −39.6 to −22.2)	<.001	−0.83	34 (82.9)
Sleep efficiency <sup>a</sup> (%)	76.2 (17.0)	84.0 (10.5)	7.8 (16.8; 2.2 to 13.3)	.007	0.54	30 (78.9)
Sleep quality (0-21)	12.2 (3.4)	6.8 (2.6)	−5.4 (4.3; −6.7 to −4.0)	<.001	−1.77	38 (90.5)
Insomnia severity (0-28)	17.3 (5.1)	9.4 (5.7)	−7.9 (6.0; −9.8 to −6.0)	<.001	−1.45	39 (92.9)
Daytime sleepiness (0-24)	10.1 (5.6)	5.4 (4.1)	−4.7 (5.2; −8.5 to −4.0)	<.001	−0.95	30 (71.4)
Sleep environment (0-39)	9.2 (6.0)	8.1 (5.3)	−1.1 (4.9; −2.6 to 0.4)	.16	−0.19	25 (59.5)
Sleep hygiene (0-52)	26.2 (8.4)	13.0 (6.6)	−13.2 (9.1; −16.0 to −9.5)	<.001	−1.72	42 (100.0)

<sup>a</sup>One outlier was removed for sleep onset for analysis (n=41), and 4 outliers were removed for sleep efficiency for analysis (n=38). All variables were assessed for normality using the Shapiro-Wilk test. Because sleep onset, daytime sleepiness, and sleep hygiene violated normality assumptions, we used the rank-based Wilcoxon signed rank test to examine the pretest-posttest differences.

**Figure 7.** Pre- and postintervention differences in sleep measurements. The red line indicates the diagonal line of equality ( $y = x$ ). Each blue dot represents an individual participant. Dots below the line indicate higher posttest values than pretest values, whereas dots above the line indicate lower posttest values than pretest values.  $r$  indicates the pre-posttest correlation for each sleep measure. Panels show (A) total sleep time, (B) sleep onset, (C) sleep efficiency, (D) sleep quality, (E) insomnia severity, (F) daytime sleepiness, (G) sleep environment, and (H) sleep hygiene.



## Discussion

### Principal Findings

To the best of our knowledge, this study was the first to assess a generative AI-based sleep chatbot intervention. Our findings support the feasibility, usability, and acceptability of this prototype AI chatbot. Improvements were found in habitual sleep patterns, subjective sleep quality, severity of insomnia, daytime sleepiness, and sleep hygiene practices after engaging with the chatbot for 2 weeks. Our study highlights the potential

of leveraging AI technologies in behavioral interventions to enhance sleep behaviors and outcomes.

Our AI chatbot was feasible in terms of implementation and user engagement. An 82.2% (88/107) sign-up rate among consenting participants suggests that the setup process, including downloading the messaging app and registering for the chatbot, was clear and easy to follow. However, the overall uptake and adherence rates were 60.7% (65/107) and 41.1% (44/107), respectively, which were slightly below expectations. This was because participants had to initiate conversations to start the intervention, an important step that we did not emphasize clearly in this study. Many waited for the chatbot to reach out, leading

to missed interactions and reducing the overall adherence. To improve future study retention, we plan to provide clearer instructions and introduce regular check-ins to increase the adherence rate. Encouragingly, participants who completed the 2-week intervention demonstrated high levels of engagement. On average, they interacted with the chatbot for 7 days, which was more frequent than the instructed minimum of once every 3 days. Throughout the study, our participants spent nearly an hour in total using the chatbot, with an average of 9 minutes per day. Although evidence on engagement with sleep chatbots is scarce, we found similar engagement rates with chatbots developed for other health behaviors. For example, one study reported an average chatbot use of 5.1 minutes per day for promoting physical activity [28]. Another study found that adolescents spent approximately 45 minutes using a health education chatbot, with sessions averaging 4 minutes [29]. Nevertheless, caution should be taken when comparing engagement rates across chatbot interventions due to variations in study focus, design, duration, and mode of delivery.

Participants reported positive experiences with the chatbot. They found the chatbot easy to use, appreciated its realistic and welcoming tone, and considered it helpful. Participants also expressed satisfaction with the chatbot's personalized and adaptive content. In general, the level of usability and acceptability of our chatbot was higher than that reported in previous research. A systematic review by Aggarwal et al [13] found that users rated chatbot interventions as having low to moderate ease of use. Additionally, fewer than 50% of users reported satisfaction with the chatbot [13]. The favorable outcomes in our study were primarily driven by the generative AI technology. LLMs are inherently adaptive, enhancing participant engagement by simulating realistic human conversations and allowing for personalized learning through reinforcement learning. As participants interact, the LLMs progressively refine responses to enhance their performance as individual data accumulate. This interactive process enables the models to update sleep recommendations in real time, which is a feature unachievable with traditional sleep interventions. Despite overall high satisfaction, several areas for improvement were identified. Some participants perceived the chatbot's responses as generic or robotic, highlighting the need for further training of LLMs to promote more natural and coherent interactions. In addition, enhancing the chatbot's memory capabilities is critical to maintain conversational continuity and deliver more personalized, contextually relevant responses.

Our findings provide preliminary evidence of sleep improvement after chatbot use. The study results echoed prior evidence on the use of conversational agents for monitoring and promoting sleep. Werner-Seidler et al [30] evaluated a 6-week app-based program that delivered CBT-I through text-based interactions with a virtual sleep therapist. Compared to the control group that received weekly sleep tips via texting, the intervention group experienced significantly greater reductions in insomnia severity and better sleep quality after the intervention and 2-month follow-up [30]. Similarly, another study developed 2 apps and randomly assigned adults with insomnia to either a control group (sleep diary only) or an intervention group (personalized recommendations from a virtual companion) [31].

After 17 days, the intervention group exhibited reduced insomnia severity, gained 48 additional minutes of sleep per night, increased sleep efficiency, and decreased wake after sleep onset ( $P < .05$ ) compared to the control group [31]. Although our chatbot achieved comparable improvements within a shorter time frame, the study findings should be interpreted with caution because all sleep outcomes were assessed via questionnaires. Discrepancies between subjective and objective sleep measurements are well documented [32,33]. One study found that self-reporting overestimated sleep duration by 64 minutes compared with actigraphy- and polysomnography-measured sleep [32]. We acknowledge that the observed sleep improvements could be subject to reporting bias. In our next trial, we will include wrist actigraphy to obtain a more accurate, comprehensive evaluation of sleep outcomes.

Several additional features of the AI chatbot support its potential for broader implementation and scalability. First, it is cost-effective. Traditional individual sleep therapy sessions such as CBT-I cost between US \$750 and US \$2500 [34]. As some insurance plans do not cover CBT-I treatments, the out-of-pocket expenses often pose financial challenges for many individuals struggling with sleep problems. In contrast, our AI chatbot is fully automated, substantially reducing the need for human resources. Unlike previous digital sleep interventions that required considerable local storage and computing power for machine learning models and data processing, our chatbot operates via LLMs hosted on cloud servers, eliminating the need for high-performance hardware. Additionally, the use of LLMs allows for seamless integration into a variety of software systems, which minimizes costs associated with app development and maintenance. Furthermore, the AI chatbot is highly accessible. Access to sleep therapy has been a long-standing issue because it is often administered by clinical specialists in person or in group settings. This issue is more notable in regions (eg, rural areas) with a shortage of trained health care providers [35]. Deploying the chatbot through a widely used commercial texting app eliminates the need for participants to learn new technology or undergo technical use training. The texting app's simple, intuitive interface enables individuals with basic smartphone skills to seek sleep support anytime and anywhere. Using a private, text-based platform can also help reduce the stigma often associated with in-person therapy, an important factor to boost engagement.

## Limitations

Despite the aforementioned advantages, our study has several limitations. First, as the quasi-experimental design cannot infer causality, a randomized controlled trial needs to be conducted to confirm the study findings. Second, we did not examine the long-term effects of the intervention. Future studies should include follow-up assessments to evaluate the sustainability of sleep improvements. Third, although participants were recruited nationwide, the study is underpowered due to its small sample size. Additionally, convenience sampling contributed to sociodemographic imbalances (eg, age, race, and educational attainment). These factors introduced selection bias and may limit the generalizability of the findings. In our next trial, we will use stratified sampling to achieve a more balanced and representative cohort. Fourth, the intervention delivery mode

may exclude certain populations, such as individuals who have insufficient smartphone technological skills, who do not own a smartphone, or who are unwilling to engage with a chatbot through texting. Finally, sleep was assessed only using self-report. Prospective work should integrate objective sleep assessments to provide a more comprehensive examination of the intervention effects.

## Conclusions

Our study demonstrates that an AI-powered, text-based chatbot is a feasible, usable, and highly acceptable intervention for

adults with poor sleep. The findings provide preliminary evidence of the association between chatbot use and improved sleep outcomes. Building on the study results, future research will refine the chatbot's design and functionality based on the limitations identified in this study. Rigorous evaluation through randomized controlled trials is necessary to validate the study findings and establish causality. If confirmed, the chatbot's affordability, engaging interface, and personalized features position it as a potential tool for large-scale implementation.

## Funding

This study was supported by New York University Rory Meyers College of Nursing and partially supported by the National Institute on Minority Health and Health Disparities (P50MD017356-4).

## Data Availability

The datasets generated or analyzed during this study are not publicly available due to the protection of participants' privacy and confidentiality but are available from the corresponding author upon reasonable request.

## Authors' Contributions

Conceptualization: XL, JL (equal)

Data curation: XL

Formal analysis: XL (lead), JL (supporting)

Funding acquisition: XL

Investigation: XL, JL (equal)

Methodology: XL, JL (equal)

Project administration: XL (lead), JL (supporting)

Resources: XL

Supervision: JL

Visualization: XL, JL (equal)

Writing—original draft: XL

Writing—review and editing: XL, JL

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Model configuration and safety settings of generative artificial intelligence.

[\[PDF File \(Adobe PDF File\), 52 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Comparison of sociodemographic and clinical characteristics between completers and noncompleters.

[\[PDF File \(Adobe PDF File\), 85 KB-Multimedia Appendix 2\]](#)

## Multimedia Appendix 3

Correlations between chatbot daily use time and study variables.

[\[PDF File \(Adobe PDF File\), 60 KB-Multimedia Appendix 3\]](#)

## References

1. Consensus Conference Panel, Watson NF, Badr MS, Belenky G, Bliwise DL, Buxton OM, et al. Recommended amount of sleep for a healthy adult: a joint consensus statement of the American Academy of Sleep Medicine and Sleep Research Society. *Sleep*. Jun 01, 2015;38(6):843-844. [\[FREE Full text\]](#) [doi: [10.5665/sleep.4716](https://doi.org/10.5665/sleep.4716)] [Medline: [26039963](https://pubmed.ncbi.nlm.nih.gov/26039963/)]
2. Lloyd-Jones DM, Allen NB, Anderson CA, Black T, Brewer LC, Foraker RE, et al. American Heart Association. Life's essential 8: updating and enhancing the American Heart Association's Construct of Cardiovascular Health: a presidential

- advisory from the American Heart Association. *Circulation*. Aug 02, 2022;146(5):e18-e43. [FREE Full text] [doi: [10.1161/CIR.0000000000001078](https://doi.org/10.1161/CIR.0000000000001078)] [Medline: [35766027](https://pubmed.ncbi.nlm.nih.gov/35766027/)]
3. Adjaye-Gbewonyo D, Ng AE, Black LI. Sleep difficulties in adults: United States, 2020. Center for Disease Control and Prevention (CDC). URL: <https://stacks.cdc.gov/view/cdc/117490> [accessed 2025-05-29]
  4. Fang H, Tu S, Sheng J, Shao A. Depression in sleep disturbance: a review on a bidirectional relationship, mechanisms and treatment. *J Cell Mol Med*. Apr 07, 2019;23(4):2324-2332. [FREE Full text] [doi: [10.1111/jcmm.14170](https://doi.org/10.1111/jcmm.14170)] [Medline: [30734486](https://pubmed.ncbi.nlm.nih.gov/30734486/)]
  5. Sofi F, Cesari F, Casini A, Macchi C, Abbate R, Gensini GF. Insomnia and risk of cardiovascular disease: a meta-analysis. *Eur J Prev Cardiol*. Jan 31, 2014;21(1):57-64. [doi: [10.1177/2047487312460020](https://doi.org/10.1177/2047487312460020)] [Medline: [22942213](https://pubmed.ncbi.nlm.nih.gov/22942213/)]
  6. Makarem N, Castro-Diehl C, St-Onge MP, Redline S, Shea S, Lloyd-Jones D, et al. Redefining cardiovascular health to include sleep: prospective associations with cardiovascular disease in the MESA sleep study. *J Am Heart Assoc*. Nov 2022;11(21):e025252. [FREE Full text] [doi: [10.1161/JAHA.122.025252](https://doi.org/10.1161/JAHA.122.025252)] [Medline: [36259552](https://pubmed.ncbi.nlm.nih.gov/36259552/)]
  7. Yang C, Yan P, Wu X, Zhang W, Cui H, Zhang L, et al. Associations of sleep with cardiometabolic risk factors and cardiovascular diseases: an umbrella review of observational and Mendelian randomization studies. *Sleep Med Rev*. Oct 2024;77:101965. [doi: [10.1016/j.smrv.2024.101965](https://doi.org/10.1016/j.smrv.2024.101965)] [Medline: [39137553](https://pubmed.ncbi.nlm.nih.gov/39137553/)]
  8. Del Brutto OH, Mera RM, Rumbela DA, Sedler MJ, Castillo PR. Poor sleep quality increases mortality risk: a population-based longitudinal prospective study in community-dwelling middle-aged and older adults. *Sleep Health*. Feb 2024;10(1):144-148. [doi: [10.1016/j.sleh.2023.10.009](https://doi.org/10.1016/j.sleh.2023.10.009)] [Medline: [38007301](https://pubmed.ncbi.nlm.nih.gov/38007301/)]
  9. Grandner MA. The cost of sleep lost: implications for health, performance, and the bottom line. *Am J Health Promot*. Sep 2018;32(7):1629-1634. [FREE Full text] [doi: [10.1177/0890117118790621a](https://doi.org/10.1177/0890117118790621a)] [Medline: [30099900](https://pubmed.ncbi.nlm.nih.gov/30099900/)]
  10. Hui SK, Grandner MA. Trouble sleeping associated with lower work performance and greater health care costs: longitudinal data from Kansas State Employee Wellness Program. *J Occup Environ Med*. Oct 2015;57(10):1031-1038. [FREE Full text] [doi: [10.1097/JOM.0000000000000534](https://doi.org/10.1097/JOM.0000000000000534)] [Medline: [26461857](https://pubmed.ncbi.nlm.nih.gov/26461857/)]
  11. Hafner M, Stepanek M, Taylor J, Troxel WM, van Stolk C. Why sleep matters-the economic costs of insufficient sleep: a cross-country comparative analysis. *Rand Health Q*. Jan 2017;6(4):11. [FREE Full text] [Medline: [28983434](https://pubmed.ncbi.nlm.nih.gov/28983434/)]
  12. Bhaskar S, Hemavathy D, Prasad S. Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities. *J Family Med Prim Care*. 2016;5(4):780-784. [FREE Full text] [doi: [10.4103/2249-4863.201153](https://doi.org/10.4103/2249-4863.201153)] [Medline: [28348990](https://pubmed.ncbi.nlm.nih.gov/28348990/)]
  13. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res*. Feb 24, 2023;25:e40789. [FREE Full text] [doi: [10.2196/40789](https://doi.org/10.2196/40789)] [Medline: [36826990](https://pubmed.ncbi.nlm.nih.gov/36826990/)]
  14. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, et al. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. *NPJ Digit Med*. Jun 23, 2023;6(1):118. [FREE Full text] [doi: [10.1038/s41746-023-00856-1](https://doi.org/10.1038/s41746-023-00856-1)] [Medline: [37353578](https://pubmed.ncbi.nlm.nih.gov/37353578/)]
  15. Lai HL, Chen CI, Liaw GH, Lu LC, Huang CY. Effectiveness of a chatbot-based internet-support-coping-caring intervention in improving sleep quality and work-related outcomes among nurses in Taiwan during the COVID-19 pandemic: a pilot study. *J Nurs Res*. Oct 01, 2025;33(5):e411. [doi: [10.1097/jnr.0000000000000691](https://doi.org/10.1097/jnr.0000000000000691)] [Medline: [40748237](https://pubmed.ncbi.nlm.nih.gov/40748237/)]
  16. Wong J, Foussat AC, Ting S, Acerbi E, van Elburg RM, Mei Chien C. A chatbot to engage parents of preterm and term infants on parental stress, parental sleep, and infant feeding: usability and feasibility study. *JMIR Pediatr Parent*. Oct 26, 2021;4(4):e30169. [FREE Full text] [doi: [10.2196/30169](https://doi.org/10.2196/30169)] [Medline: [34544679](https://pubmed.ncbi.nlm.nih.gov/34544679/)]
  17. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. *Behav Sleep Med*. Dec 28, 2011;10(1):6-24. [FREE Full text] [doi: [10.1080/15402002.2012.636266](https://doi.org/10.1080/15402002.2012.636266)] [Medline: [22250775](https://pubmed.ncbi.nlm.nih.gov/22250775/)]
  18. Holmes S, Moorhead A, Bond R, Zheng H, Coates V, Mctear M. Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces? In: *Proceedings of the 31st European Conference on Cognitive Ergonomics*. 2019. Presented at: ECCE '19; September 10-13, 2019:207-214; London, United Kingdom. URL: <https://dl.acm.org/doi/10.1145/3335082.3335094> [doi: [10.1145/3335082.3335094](https://doi.org/10.1145/3335082.3335094)]
  19. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. *Usability Evaluation In Industry*. Boca Raton, FL. CRC Press; 1996:189-194.
  20. Lewis JR, Sauro J. Item benchmarks for the System Usability Scale. *J Usability Stud*. 2018;13(3):158-167. [FREE Full text] [doi: [10.1145/3335082.3335094](https://doi.org/10.1145/3335082.3335094)]
  21. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. *Eval Program Plann*. Jan 1979;2(3):197-207. [doi: [10.1016/0149-7189\(79\)90094-6](https://doi.org/10.1016/0149-7189(79)90094-6)] [Medline: [10245370](https://pubmed.ncbi.nlm.nih.gov/10245370/)]
  22. Buysse DJ, Reynolds 3rd CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psychiatry Res*. May 1989;28(2):193-213. [doi: [10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4)] [Medline: [2748771](https://pubmed.ncbi.nlm.nih.gov/2748771/)]
  23. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med*. Jul 2001;2(4):297-307. [doi: [10.1016/s1389-9457\(00\)00065-4](https://doi.org/10.1016/s1389-9457(00)00065-4)] [Medline: [11438246](https://pubmed.ncbi.nlm.nih.gov/11438246/)]
  24. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*. Dec 1991;14(6):540-545. [doi: [10.1093/sleep/14.6.540](https://doi.org/10.1093/sleep/14.6.540)] [Medline: [1798888](https://pubmed.ncbi.nlm.nih.gov/1798888/)]



25. Mastin DF, Bryson J, Corwyn R. Assessment of sleep hygiene using the Sleep Hygiene Index. *J Behav Med*. Jun 24, 2006;29(3):223-227. [doi: [10.1007/s10865-006-9047-6](https://doi.org/10.1007/s10865-006-9047-6)] [Medline: [16557353](https://pubmed.ncbi.nlm.nih.gov/16557353/)]
26. Grandner MA, Valencia DY, Seixas AA, Olivieri K, Gallagher RA, Killgore WD, et al. Development and initial validation of the Assessment of Sleep Environment (ASE): describing and quantifying the impact of subjective environmental factors on sleep. *Int J Environ Res Public Health*. Oct 20, 2022;19(20):13599. [FREE Full text] [doi: [10.3390/ijerph192013599](https://doi.org/10.3390/ijerph192013599)] [Medline: [36294179](https://pubmed.ncbi.nlm.nih.gov/36294179/)]
27. R Core Team. The R project for statistical computing. R Foundation for Statistical Computing. URL: <https://www.r-project.org/> [accessed 2025-05-29]
28. To QG, Green C, Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning-based physical activity chatbot: quasi-experimental study. *JMIR Mhealth Uhealth*. Nov 26, 2021;9(11):e28577. [FREE Full text] [doi: [10.2196/28577](https://doi.org/10.2196/28577)] [Medline: [34842552](https://pubmed.ncbi.nlm.nih.gov/34842552/)]
29. Crutzen R, Peters GJ, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. *J Adolesc Health*. May 2011;48(5):514-519. [doi: [10.1016/j.jadohealth.2010.09.002](https://doi.org/10.1016/j.jadohealth.2010.09.002)] [Medline: [21501812](https://pubmed.ncbi.nlm.nih.gov/21501812/)]
30. Werner-Seidler A, Li SH, Spanos S, Johnston L, O'Dea B, Torok M, et al. The effects of a sleep-focused smartphone application on insomnia and depressive symptoms: a randomised controlled trial and mediation analysis. *J Child Psychol Psychiatry*. Sep 29, 2023;64(9):1324-1335. [FREE Full text] [doi: [10.1111/jcpp.13795](https://doi.org/10.1111/jcpp.13795)] [Medline: [36991537](https://pubmed.ncbi.nlm.nih.gov/36991537/)]
31. Philip P, Dupuy L, Sagaspe P, de Sevin E, Auriacombe M, Taillard J, et al. Efficacy of a smartphone-based virtual companion to treat insomniac complaints in the general population: sleep diary monitoring versus an internet autonomous intervention. *J Clin Med*. Jul 28, 2022;11(15):4387. [FREE Full text] [doi: [10.3390/jcm11154387](https://doi.org/10.3390/jcm11154387)] [Medline: [35956005](https://pubmed.ncbi.nlm.nih.gov/35956005/)]
32. Jackson CL, Patel SR, Jackson II WB, Lutsey PL, Redline S. Agreement between self-reported and objectively measured sleep duration among white, black, Hispanic, and Chinese adults in the United States: multi-ethnic study of atherosclerosis. *Sleep*. Jun 01, 2018;41(6):zsy057. [FREE Full text] [doi: [10.1093/sleep/zsy057](https://doi.org/10.1093/sleep/zsy057)] [Medline: [29701831](https://pubmed.ncbi.nlm.nih.gov/29701831/)]
33. Jackson CL, Ward JB, Johnson DA, Sims M, Wilson J, Redline S. Concordance between self-reported and actigraphy-assessed sleep duration among African-American adults: findings from the Jackson Heart Sleep Study. *Sleep*. Mar 12, 2020;43(3):zsz246. [FREE Full text] [doi: [10.1093/sleep/zsz246](https://doi.org/10.1093/sleep/zsz246)] [Medline: [31616945](https://pubmed.ncbi.nlm.nih.gov/31616945/)]
34. Daley K. Is cognitive behavioral therapy for insomnia worth the cost: find your BASE. BASE Cognitive Behavioral. URL: <https://www.findyourbase.com/blog/2019/3/is-cognitive-behavioral-therapy-for-insomnia-worth-the-cost> [accessed 2024-06-17]
35. Morin CM. Cognitive behavioural therapy for insomnia (CBTi): from randomized controlled trials to practice guidelines to implementation in clinical practice. *J Sleep Res*. Apr 12, 2020;29(2):e13017. [doi: [10.1111/jsr.13017](https://doi.org/10.1111/jsr.13017)] [Medline: [32164036](https://pubmed.ncbi.nlm.nih.gov/32164036/)]

## Abbreviations

**AI:** artificial intelligence  
**CBT-I:** cognitive behavioral therapy for insomnia  
**CSQ:** Client Satisfaction Questionnaire  
**CUQ:** Chatbot Usability Questionnaire  
**LLM:** large language model  
**MD:** mean difference

*Edited by A Schwartz, M Balcarras; submitted 12.Sep.2025; peer-reviewed by M Al-Agil, Y Chen, G Xu, J DOng; comments to author 27.Oct.2025; accepted 25.Nov.2025; published 03.Feb.2026*

### *Please cite as:*

Liu X, Liu J

*Assessing the Feasibility, Usability, Acceptability, and Efficacy of an AI Chatbot for Sleep Promotion: Quasi-Experimental Study*  
*JMIR Form Res* 2026;10:e84023

URL: <https://formative.jmir.org/2026/1/e84023>

doi: [10.2196/84023](https://doi.org/10.2196/84023)

PMID:

©Xiaoyue Liu, Jingchen Liu. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 03.Feb.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.