Enhancing physical activity through a relational artificial intelligence chatbot: A feasibility and usability study

Yoo Jung Oh[1] (ID), Kai-Hui Liang[2], Diane Dagyong Kim[3], Xuanming Zhang[2],
Zhou Yu[2], Yoshimi Fukuoka[4] (ID) and Jingwen Zhang[3] (ID)

Abstract

Objective: This study presents a pilot randomized controlled trial to assess the usability, feasibility, and initial efficacy of a mobile app-based relational artificial intelligence (AI) chatbot (Exerbot) intervention for increasing physical activity behavior.

Methods: The study was conducted over a 1-week period, during which participants were randomized to either converse with a baseline chatbot without relational capacity (control group) or a relational chatbot using social relational communication strategies. Objectively measured physical activity data were collected using smartphone pedometers.

Results: The study was feasible in enrolling a sample of 36 participants and with a 94% retention rate after 1 week. Daily engagement rate with the AI chatbot reached over 88% across the groups. Findings revealed that the control group experienced a significant decrease in steps on the final day, whereas the group interacting with the relational chatbot maintained their step counts throughout the study period. Importantly, individuals who engaged with the relational chatbot reported a stronger social bond with the chatbot compared to those in the control group.

Conclusions: Leveraging AI chatbot and the relationship-building capabilities of AI holds promise in the development of cost-effective, accessible, and sustainable behavior change interventions. This approach may benefit individuals with limited access to conventional in-person behavior interventions.

Clinical trial registrations: ClinicalTrials.gov; NCT05794308; https://clinicaltrials.gov/ct2/show/NCT05794308.

Introduction

Physical activity promotion has faced challenges in achieving population-wide reach and long-term behavior maintenance due to the lack of access to and high cost of in-person counseling and interventions.[1] Despite efforts to raise awareness and involve communities for physical activity programs, 80% of US adults currently fail to meet recommended physical activity guidelines, contributing to increased health risks and healthcare expenditures.[2,3] The limited accessibility of in-person counseling and programs

[1]Department of Communication, Michigan State University, East Lansing, Michigan, USA
[2]Department of Computer Science, Columbia University, New York, USA
[3]Department of Communication, University of California, Davis, California, USA
[4]Department of Physiological Nursing, University of California, San Francisco, California, USA

Corresponding author:
Yoo Jung Oh, Department of Communication, Michigan State University, 404 Wilson Road, East Lansing, Michigan 48824, USA. E-mail: ohyoojun@msu.edu

has necessitated cost-effective, scalable, and sustainable intervention alternatives.[4–6]

Recent advancements in artificial intelligence (AI), natural language processing, and machine learning have enabled AI chatbot-based physical activity interventions to emerge as a promising solution. AI chatbots are intelligent dialogue systems designed to simulate human communication through text, speech, or both. Chatbots can be incorporated into mobile applications, providing round-the-clock communication with users.[7,8] Chatbots have the potential to offer cost-effective, accessible, sustainable, and personalized health interventions tailored to individual needs.[9]

Research has suggested AI chatbots can provide dependable, humanlike interactions in healthcare settings[6,10,11] and AI chatbots can promote lifestyle modifications, such as physical activity and healthy diets.[12–14] For example, To and colleagues[14] found a machine learning-based chatbot integrated with Fitbit increased users' average daily steps by 627 after 6 weeks of intervention.

Despite evidence suggesting AI chatbot's potential efficacy, little attention has been paid to evaluating human perceptions toward AI chatbots in the context of health behavior interventions.[13] Specifically, whether human users can develop trust and positive relational perceptions toward AI systems is greatly relevant in determining the acceptance and quality of AI chatbot interventions, which will gain increasing prominence in the next decade. Thus, it is crucial to investigate whether AI chatbots can elicit and establish positive relationships with users, as health outcomes often depend on long-term patient–provider relationships (e.g., counselors).[15,16]

Considering the limited research on the effectiveness of AI-chatbot-based physical activity interventions and human–chatbot relationships in healthcare contexts, this study presents a "relational AI chatbot," named Exerbot, using relationship-building strategies to enhance an evidence-based physical activity intervention via a mobile application. This paper reports the feasibility, usability, and initial efficacy of a 1-week chatbot intervention on increasing objectively assessed physical activity data.

Literature review

Artificial intelligence (AI) has emerged as an innovative and essential technology that reforms how we analyze information and improve decision-making. The recent advancements in AI and its applications have shown significant progress in healthcare. For instance, AI-driven chatbot-based interventions have been used as viable strategies for enhancing public health communication. AI chatbots are conversational agents that use AI and machine learning technologies to mimic human interactions through engaging in natural language interactions with users.[7] These chatbots offer cost-effective, accessible,

sustainable, and personalized health interventions that are tailored to individual needs,[9] promoting positive health behavior change.[12,13] Oh and colleagues[13] discovered that out of seven studies on physical activity outcomes, five demonstrated statistically significant improvements. For instance, in Maher et al.,[9] participants conversing with an AI chatbot Paola while using a step tracker increased their physical activity by an average of 109.8 minutes per week at 12 weeks. Yet, while most studies report improvement in health behavior change in the short-term, or evaluate outcomes immediately before and after interventions, a systematic review by Han and colleagues[17] finds that there is limited evidence for the feasibility and acceptability of chatbots in promoting nutrition and physical activity behaviors, particularly regarding adherence. This highlights the challenge of maintaining behavioral changes over time and points to a need for further research on how chatbots can leverage behavioral theories and persuasive strategies to enhance sustained engagement.

Despite the potential benefits of AI chatbots, further challenges and limitations exist. For instance, previous studies[18,19] discuss that the existing digital therapeutic interventions with didactic components have encountered several challenges, including relatively low adherence, unsustainability, and inflexibility. In addition, when it comes to the feasibility, acceptability, and usability of AI chatbots in various healthcare applications, findings are mixed. In a systematic review, Aggarwal and colleagues[12] found that evidence on the safety and feasibility of chatbots was limited, with only a small percentage of studies reporting on these metrics. In addition, the acceptability of chatbots was found to be less than ideal, with fewer than 50% of participants expressing satisfaction.[12] This aligns with previous systematic reviews that highlight the challenges in achieving high levels of user acceptance and engagement over time.[20,21] The usability of AI chatbots also varied across studies, with some reporting positive experiences in terms of efficient support and content reliability, while others noted an ambiguity in the content, technical difficulties, such as limited responsiveness and failure to understand user inputs, and a need for improvement in quality of recommendation provided by AI chatbots for feasible implementation.[12]

Moreover, only a few studies have investigated the specific design considerations that could improve relational qualities. For example, previous works have shown that digital emphatic agents are perceived as more caring, sympathetic and trustworthy than agents without empathic abilities,[22] and that changing the initiative to be more proactive, such as through small talk, seems to be welcomed and contributes to building close relationships by users.[23,24] More rigorous research needs to be conducted to explore the development of AI-based chatbots that are designed to foster a sense of warmth, empathy, and personal connection with users. By creating chatbots that are not only

functionally capable, but also emotionally and socially relational, researchers may be able to address the current limitations in usability and acceptability, improving the overall effectiveness of the conversational AI systems in healthcare.

Therefore, the primary goal of this study is to examine the feasibility of AI chatbot-based physical activity counseling and see whether users find this usable. Specifically, we explore the relational chatbot's influence on behavioral engagement and usability perceptions. Research suggests that user engagement and satisfaction can be enhanced by incorporating humanlike qualities into computer systems or interfaces.[25] As a result, it is essential to investigate how chatbot's relational behaviors may impact user engagement outcomes when delivered via a mobile app, combined with the availability of behavioral data.

Additionally, in chatbot-based health interventions, chatbots' relational behaviors can increase the alliance between chatbots and users, helping users to adhere to chatbot's health advice. Cumulative evidence in clinical psychology, counseling, and coaching has acknowledged the positive impact of strong physician–patient relationships on psychotherapeutic outcomes.[26] When patients and physicians establish collaborative relationships, patients are more inclined to perceive treatment value and feel empowered toward behavior change and adherence.[27–29] Physician–patient relationships, defined by mutual liking, trust, and belief in achieving desired outcomes, have been consistently linked to critical treatment outcomes like patient adherence.[27,30–32]

Based on this idea, previous research has utilized physicians' socio-emotional and relational communicative behaviors, such as empathy, to incorporate into chatbots and promote human–chatbot relationships.[15] For example, studies have demonstrated the effectiveness of conversational agents' verbal relational behaviors, such as social dialogue,[22] self-disclosure,[33] empathy,[34] meta-relational communication,[35] and humor,[15] in fostering positive relationships in health and well-being contexts. It has been shown that chatbots employing relational behaviors, such as social support and self-disclosure, are more likely to establish positive human–chatbot relationships such as therapeutic alliance, enhance continued engagement, and increase physical activity.[36,37] Therefore, as a secondary objective, this study examines the initial efficacy of an AI chatbot-based intervention, focusing on the potential of the relational chatbot in enhancing human–AI relationship perceptions as well as physical activity outcomes.

Methods

Chatbot (Exerbot) development and app design

The chatbot, named Exerbot, was developed using a hybrid approach, combining rule-based methods and a fine-tuned language model. Its structure consisted of two main components: a topic proposer and a dynamic response generator.

The topic proposer used predefined templates designed for physical activity intervention. The dynamic response generator is composed of a generation-based model and a retrieval-based question handler. The generation-based model addresses users' general inquiries and acknowledges their responses. To generate plausible and natural responses, it was built with a conversational generation model, Blender,[38] which is trained on large amounts of human–human conversation data and is adjusted to learn several conversation skills from humans. To handle topics relevant to physical activity intervention, the model was initially trained on large-scale human–human conversation data and further fine-tuned using transcripts from 105 in-person physical activity counseling sessions collected from the mobile phone-based physical activity education program (mPED) study.[39–41] The question handler is a retrieval-based question-answering module built with a custom database to address the most frequently asked questions related to physical activities. During each dialog turn, the chatbot generated a response using the dynamic response generator and concatenated the response with the template from the topic proposer, ensuring both contextual relevance and adherence to physical activity intervention strategies (see[42] for a detailed description of the dataset). The physical activity interventions in the topic proposer were designed based on Social Cognitive Theory.[43]

Then, the Exerbot app was designed for research purposes to offer educational counseling sessions to encourage individuals to enhance and sustain their physical activity levels. As illustrated in Figure 1, the Exerbot presents an instruction page (left) that provides general information about the study, allowing participants to understand and explore the app's features. The chatroom interface (middle) provides a space for participants to engage in daily conversations with Exerbot. Lastly, the app also features a dashboard displaying daily and weekly step counts, as shown in Figure 1 (right). The app interface and chatbot were prototyped by researchers, with a professional app developer handling the front-end and back-end programming, as well as the deployment of the app. Exerbot app supported in Apple's iOS to reach iPhone users.

Upon installation of the app, Exerbot greets the participant and commences the initial educational session. Over the course of 7 days, the app conducts daily sessions on physical activity while tracking the user's progress via the smartphone's built-in pedometer. Exerbot offers visualizations of the past 7 days' step counts and encourages participants to consistently monitor their progress through daily reminders. Each time Exerbot initiates a conversation or sends a reminder, a push notification appears on the participant's smartphone status bar and lock screen (e.g., "Time for a quick check-in! Go to Stats tab to track your daily step progress"), prompting them to engage with Exerbot or assess their physical activity status.
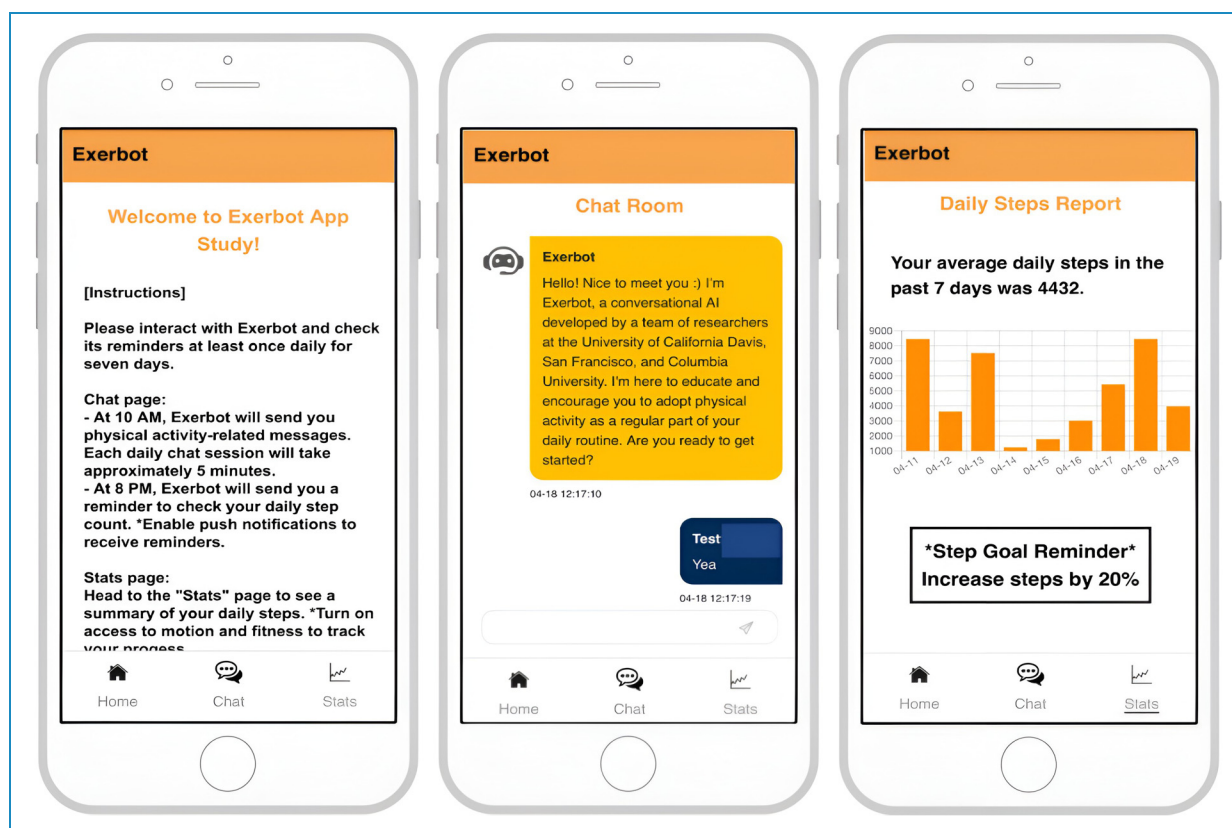
Figure 1. Exerbot app: Home page and study overview (left), chatroom with Exerbot (middle), and daily and weekly step report (right).

Study design and sample

A pilot randomized controlled trial (RCT) with two groups was conducted. Data were collected via the Exerbot app on participants' smartphones and from pre- and postsurveys. Eligibility criteria included being: (1) 18 years or older, (2) able to read and speak English, (3) residing in the United States, (4) iPhone users, and (5) inactive (i.e., not meeting 10,000 daily steps). Eligible participants who completed the presurvey were invited for the app-based intervention. They received information about study goals and procedures and, if willing to participate, were given a study link to enroll. After enrollment, they were instructed to download and install the Exerbot app. Upon signing up, participants enabled pedometer access and allowed push notifications. Participants were randomly allocated to either the relational chatbot condition or nonrelational (control) chatbot condition version of the app. Researchers and participants were blinded to treatment allocation. The Exerbot initiated the first educational session, and participants viewed a graph of their past 7-day step counts. The intervention lasted 7 days, with a postsurvey on the final day. Daily conversations began at 10AM, and progress reminders were sent at 8PM.

This study received approval from the authors' Institutional Review Board (IRB; approval number: 2033031-1), ensuring adherence to ethical guidelines and standards in academic research. Participants provided electronic informed consent before they started the eligibility screening survey. They were presented with a comprehensive consent page detailing the study's purpose, procedures, potential risks and benefits, and the voluntary nature of participation. Consent was obtained by requiring participants to actively click a button indicating their agreement after they read the full consent form and before they proceeded to the eligibility screening survey. This process ensured informed and voluntary participation while maintaining participant anonymity, as no personally identifiable information was collected.

Intervention components and randomization

Daily conversation modules

On day 1, Exerbot gave an overview of the study, talked about the benefits of physical activity, and set a personalized physical activity goal. The short-term goal was set to increase step counts by 20% compared to the participant's average step counts of the previous 7 days, which were accessed via the smartphone's pedometer activity history upon app installation. The long-term goal was set to maintain 10,000 steps per day, which approximates the World Health Organization's physical activity

recommendations.[44] On day 2, Exerbot identified the participant's barriers to physical activity and ways to overcome them. On day 3, Exerbot talked about establishing a physical activity routine and coordinated with participants to create a personalized physical activity routine. On day 4, Exerbot talked about social support in physical activity and asked participants to identify people who can check in and offer support. On day 5, Exerbot provided information on how to prevent relapse and keep oneself motivated. On day 6, Exerbot provided education sessions about diet and weight maintenance. Lastly, on day 7, a discussion about safety was provided. In the closing statement, Exerbot informed participants to complete the postsurvey.

Manipulation of relational and nonrelational chatbot conditions

Upon signing in, participants were allocated to one of two chatbot conditions: relational or nonrelational (control). In the relational chatbot condition, five relational cues were incorporated into the daily chat sessions provided by Exerbot. Conversely, the nonrelational (control) chatbot condition involved daily educational sessions without any inclusion of relational cues. Specific expressions for each relational cue were extracted from mPED trial counseling sessions as well as dialogue examples provided in the previous literature.[15,36,37,45]

The five relational cues are described as follows. Social dialogue included informal greetings (e.g., "Good morning! How are you? Are you ready to get started for today's chat session?") and farewells (e.g., "Keep up the good work! Talk to you again tomorrow!"), which were designed according to the day of the intervention to provide feelings of continuity. Self-disclosure included responses that disclosed information about the chatbot's experiences and corresponding feelings related to the daily educational content (e.g., Social support-related content; "For example, I have a friend named BodBot who offers me support and I'm grateful for that"). Empathy included emotional responses such as "I understand how frustrating it is to…" when the user disclosed information relevant to potential barriers or struggles in physical activity. Meta-relational communication included expressions to check on the state of the conversation (e.g., "How do you feel our conversation is going so far?"). Lastly, humor included exercise-related jokes (e.g., "Do you know what is a sloth's favorite form of exercise? Running late!"). To ensure agent variability, multiple expressions for each relational cue were implemented in the chatbot.

Outcome measures

To evaluate feasibility and usability as our primary objective, (1) retention, (2) chatbot usability, and (3) user engagement were measured. Retention was measured by the number of participants who completed the postsurvey compared with those who finished the baseline questionnaire. Chatbot usability and user satisfaction were measured using technology-based intervention metrics developed in previous studies.[46–48] Chatbot usability was measured by a seven-item scale ($M = 3.97$, $SD = 0.51$, $\alpha = 0.70$) (e.g., "Chatbot responses were useful, appropriate, and informative."). User satisfaction was rated on a five-point scale ranging from "very dissatisfied" to "very satisfied" (single-item; $M = 3.71$, $SD = 0.84$) (single-item; "In general, are you satisfied with the chatbot?").

User engagement was measured through objective measures such as total days of completed chat sessions and participants' daily average message length (i.e., word count). Additionally, we investigated the level of cognitive engagement to understand user interaction with the chatbot. Specifically, we explored how participants' language use varied when interacting with two different types of Exerbots. Linguistic Inquiry and Word Count (LIWC) was used to quantify various dimensions of language use and analyze these differences.[49] Specifically, we focused on the use of pronouns, prepositions, auxiliary verbs, adverbs, cognitive processing words, causations, motion words, and expressive punctuations, such as exclamation marks and question marks.

Individual's writing styles and syntactic choices can reveal the depth and complexity of the cognitive processes. According to Tausczik and Pennebaker,[50] complex language use is characterized by prepositions, cognitive mechanisms, and words greater than six letters, which captures cognitive depth. For example, prepositions allow for more concrete information about a topic, enhancing the clarity. Similarly, the use of causal words (such as "how," "because," "why") and insight words (such as "think," "know," "feel"), which are subcategories of cognitive mechanisms, demonstrate the active process of reappraisal.[50] Furthermore, nouns and other modifiers such as adverbs and adjectives are used more to form sentences with high intensity, leading to clearer message perception.[51] As such, the use of complex sentences can be indicative of the communicator's willingness to interact and engage with the other entity. In sum, analyzing various dimensions of users' language use can reveal how different chatbot interaction styles influence users' language output, thereby providing insights into the preferences and tendencies of users.

To measure human–chatbot relationships as our secondary objective, therapeutic alliance (TA) was used for relationship evaluation. TA consists of three subcomponents: bond, goal, and task. A strong TA in human–chatbot interaction will reflect a well-functioning relationship where the human user has positive attachments toward the chatbot, endorses the mutual goal, and accepts the task.[15,52,53] Example items for bond, goal, and task were "I believe the chatbot likes me," "the chatbot and I are working

towards mutually agreed upon goals," and "what I am doing in this physical activity conversation gives me new ways of looking at my problem," respectively (combined; $M = 3.99$, $SD = 0.52$, $\alpha = 0.79$).

Lastly, physical activity outcomes were assessed through both objective and subjective metrics:[54–56] (1) daily and weekly step counts obtained from a smartphone (2) daily and weekly step goal achievement (i.e., 20% increase from baseline), and (3) poststudy physical activity intention, measured by a three-item scale ($M = 3.86$, $SD = 0.97$, $\alpha = 0.88$). Step count information was obtained from participants' smartphones.

Statistical analyses

Feasibility, usability, and participants' sociodemographic characteristics were analyzed descriptively, using means, standard deviations, frequencies, and percentages. To evaluate the differential effects of relational and nonrelational (control) chatbots on therapeutic alliance (TA) and physical activity outcomes such as step goal achievements, independent *t*-tests were conducted.

In addition, to assess changes in physical activity during the intervention period, step count data were collected using pedometers over a 14-day period (7 days preintervention and 7 days during the intervention). The average of the 7 days preceding the intervention was calculated to establish a baseline step count. To analyze the data, we employed a linear mixed-effects model, with the baseline and the subsequent seven timepoints recorded each day as dependent variables. The model included participants as random intercepts to account for within-participant correlations of repeated measures. Fixed effects comprised chatbot conditions, seven daily time indicators, and chatbot condition-by-time indicator interactions. The model was fitted using restricted maximum likelihood (REML) estimation. Analyses were conducted using the lme4 package in R, with a total of 285 observations from 36 participants included in the model. All statistical analyses were conducted using SPSS version 23.0 and R software. Statistical significance was defined as $P < .05$ for all tests.

Results

Recruitment and retention

As summarized in Figure 2, initially, 222 participants signed up for the study. They were informed that this was a two-part study, consisting of an eligibility survey and an app-based intervention. After eligibility screening, 165 (74.3%) were deemed eligible and invited to join the app-based intervention. Out of 38 who signed up to engage in the app-based intervention, two did not register in the mobile app, resulting in 36 enrolled participants. Twenty-one were randomly assigned to the relational

chatbot condition, while 15 were assigned to the nonrelational (control) chatbot condition. Retention after the 1-week intervention period was 94.4%.

Participant characteristics

On average, participants were aged 21.0 years ($SD = 6.27$), and 75.0% were women. 58.3% were Asian or Pacific Islander, 19.4% were Hispanic, 8.3% were White, and 13.9% were mixed race. The majority of participants (72.2%) had experience using chatbots before. Baseline daily steps were 4059.40 ($SD = 3838.43$).

Primary outcomes: engagement and usability

To measure the level of engagement across the 7-day intervention, participants in the relational chatbot condition completed daily chat sessions on 6.19 days ($SD = 1.17$), whereas participants in the control condition completed the daily chat sessions on 6.73 days ($SD = 0.46$). Both conditions demonstrated over 88% daily engagement, indicating that intervention was feasible. Although only marginally significant, participants in the relational chatbot condition had a total average word count of 192.90 ($SD = 109.01$), while those in the control condition had a total average word count of 138.73; $t(30.87) = 1.97$, $P = .058$. This suggests that participants who interacted with a chatbot that engages in relational behaviors tended to respond in lengthier messages.

Overall, participants deemed the Exerbot to be usable ($M = 3.97$, $SD = 0.51$), and were satisfied with the system ($M = 3.71$, $SD = 0.84$), but were relatively reluctant to continue using the system after the intervention period ($M = 2.47$, $SD = 0.93$). Usability, satisfaction, and continued engagement intention did not differ by the two conditions ($P > .05$).

As presented in Table 1, findings revealed significant differences between the two chatbot conditions. Participants who interacted with the relational chatbot were more likely to use first-person pronouns ($P = .030$), prepositions ($P = .039$), auxiliary verbs ($P = .005$), adverbs ($P = .002$), and question marks ($P = .013$) compared to those who interacted with the nonrelational (control) chatbot. In addition, the participants used more cognitive processes ($P = .022$), insights ($P = .035$), and causation words ($P < .001$). While showing marginal significance ($P = .058$), participants in the relational condition showed higher frequency in total word count as well as average words per sentence. In addition, they used more clout (i.e., language of leadership, status), and authentic (i.e., perceived honesty, genuineness) words.

In contrast, participants interacting with the nonrelational (control) chatbot exhibited a higher frequency of negations ($P < 0.001$) and assent ($P = .043$) compared to those who interacted with the relational chatbot. Although
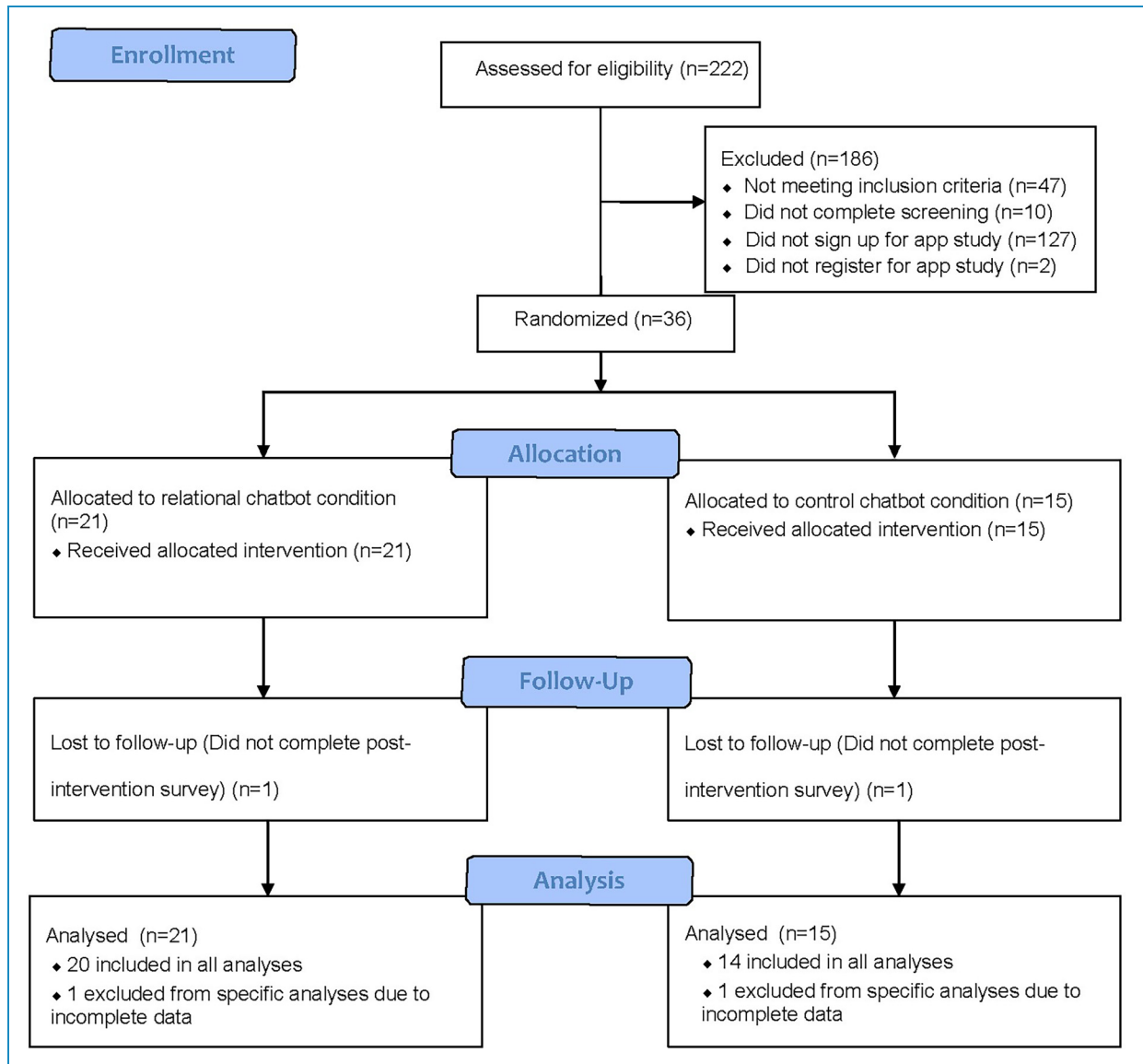
Figure 2. Consort flow diagram of study participants.

the differences were only marginal ($P < .1$), participants in the control condition tended to use more affect-based words, express positive emotions, and employ informal language.

Secondary outcomes: human–chatbot relationship and health behavior

Regarding human–chatbot relationship outcomes of the relational and nonrelational (control) chatbot conditions, there was no significant difference in the overall therapeutic alliance. However, a marginally significant difference emerged in the "bond" dimension ($P = .081$). Participants interacting with the relational chatbot perceived a stronger social and emotional bond with the chatbot ($M = 4.49$, SD

$= 0.49$) compared to those engaging with the control chatbot ($M = 4.18$, $SD = 1.09$).

Throughout the intervention period, participants increased their steps, with the relational chatbot condition showing an average increase of 392.62 steps from a baseline of 3882.61 ($SD = 1740.50$), and the nonrelational (control) chatbot condition an increase of 494.01 steps from a baseline of 4306.93 ($SD = 2306.64$). Despite the control chatbot condition showing a greater increase in steps than the relational chatbot condition, this difference was not found to be statistically significant. Furthermore, participants in the relational chatbot condition, on average, achieved their step goals (i.e., a 20% increase from baseline steps) on 2.7 days per week, while those in the control condition met their step goals on 3.14 days

Table 1. Comparison of linguistic characteristics of user responses between relational and nonrelational (control) chatbot conditions.

LIWC category	Relational (n = 21)	Nonrelational (n = 15)		
	Mean (SD)	Mean (SD)	t	P
Word count	193.26 (109.35)	138.73 (53.89)	−1.975	.057#
Clout (language of leadership, status)	45.13 (5.35)	42.27 (3.49)	−1.804	.080#
Authentic (perceived honesty, genuineness)	28.00 (12.02)	21.41 (5.35)	−1.835	.075#
Words per sentence	3.02 (1.36)	2.26 (0.95)	−1.861	.071#
1st person pronoun	3.69 (2.27)	2.11 (1.75)	−2.259	.030*
Prepositions	3.54 (2.60)	1.96 (1.41)	−2.149	.039*
Auxiliary verbs	3.61 (2.58)	1.70 (1.07)	−3.034	.005**
Adverbs	4.83 (3.07)	2.01 (2.07)	−3.289	.002**
Negations	3.79 (2.67)	10.72 (2.84)	7.471	<.001***
Verbs	9.25 (4.31)	6.58 (4.32)	−1.829	.076#
Numbers	1.13 (0.68)	0.78 (0.47)	−1.854	.072#
Affect	29.29 (8.69)	34.76 (7.99)	1.925	.063#
Positive emotions	28.88 (8.92)	34.33 (7.96)	1.886	.068#
Cognitive processes	13.42 (6.13)	8.82 (4.94)	−2.401	.022*
Insight	1.95 (1.26)	1.1 (0.92)	−2.193	.035*
Causation	3.08 (2.79)	0.62 (0.66)	−3.890	<.001***
Drives (we, our, work, us)	13.07 (3.59)	10.28 (3.58)	−2.297	.028*
Reward (win, gain, benefit)	7.96 (3.30)	5.99 (3.10)	−1.182	.079#
Motions (go, come, went, came)	2.40 (1.44)	1.26 (0.97)	−2.670	.012*
Informal	33.92 (13.31)	41.95 (13.21)	1.791	.082#
Assent (yes, ok)	31.87 (12.39)	40.98 (13.43)	2.100	.043*
Question mark	2.53 (3.19)	0.55 (0.95)	−2.688	.013*

Note. #P < .1, *P < .05, **P < .01, ***P < .001.

per week. A marginal effect of the relational chatbot condition on participants' self-reported PA intention was observed. Participants interacting with the relational chatbot reported higher levels of PA intention ($M = 4.10$, $SD = 0.97$) compared to those engaging with the control chatbot ($M = 3.52$, $SD = 0.89$), $P = .088$.

To capture a more nuanced trend across the time points on daily step counts, a linear mixed-effects model was conducted to examine the impact of timepoint and relational condition on the number of steps taken, with a random intercept included for subjects to account for individual variability. As presented in Table 2, the analysis revealed a significant main effect of timepoint (estimate = −243.98, $SE = 118.47$, $t(248.33) = −2.059$, $P = .041$, 95% CI: [−476.2, −11.8]), indicating that the number of steps decreased over time. The main effect of the relational

condition was marginally significant (estimate = −1551.47, SE = 886.00, t(80.22) = −1.751, P = .084, 95% CI: [−3288.0, 185.1]), which indicated that those in the relational condition took fewer steps compared to those in the control condition. Interestingly, there was a marginally significant interaction between timepoint and relational condition (estimate = 301.08, SE = 154.42, t(247.91) = 1.950, P = .052, 95% CI: [−1.6, 603.7]). This suggests that the decline in steps over time was less pronounced in the relational condition, highlighting a potential mitigating effect of the relational chatbot on the reduction in physical activity levels over time.

Figure 3 further showcases the changes in daily steps throughout the study period. For the initial six days, there was no significant difference in daily step outcomes between the relational and nonrelational (control) chatbot conditions; in fact, participants in the nonrelational (control) chatbot condition exhibited higher daily steps. Intriguingly, on the final day of the intervention, a significant decrease in steps was observed in the nonrelational (control) chatbot condition.

Adverse events

No adverse events related to study participation were experienced.

Discussion

Principal findings

This study investigated the feasibility, usability, and efficacy of an AI-powered relational chatbot designed to promote physical activity. In the study, a high retention rate of 94.4% was achieved, with only 2 out of 36 participants not completing the postsurvey. User engagement, measured by the daily chat completion rate, was high for both conditions, suggesting good feasibility outcomes. However, considering that comparable mHealth studies employing chatbots have reported retention rates of approximately 90% over 12-week study periods,[9,57] it is crucial for future research to incorporate strategies that maintain participant engagement throughout the study. One possible strategy to increase long-term engagement is to personalize interactions with users. For example, just-in-time adaptive intervention (JITAI) allows for combining real-time data from built-in pedometers and other smartphone sensors to examine specific situations in which participants are most likely to react positively to notifications and tailor the timing, content, and frequency of the intervention to each person.[55,58]

As for linguistic characteristics of user's responses to Exerbot, participants in the relational chatbot condition demonstrated a greater use of total word counts as well as linguistic markers such as first-person pronouns, prepositions, auxiliary verbs, adverbs, and question marks, which can be linguistic markers of personal and involved style of communication. This suggests that relational chatbots can encourage a sense of personalization and enhance user engagement that promotes richer and more dynamic interaction with the chatbot. These findings align with established principles regarding the value of building relational interactions to optimize user experiences. For instance, Bickmore and Picard[15] highlighted how the incorporation of relational elements, such as the use of conversational cues and personalized language, can promote a stronger sense of engagement and connection between users and conversational agents. Moreover, the greater usage of cognition-related words, such as those reflecting cognitive processes, insights, and causation, by participants in the relational chatbot condition indicates a more analytical and explanatory style of communication. This suggests that the relational chatbot prompted users to engage in a more complex, engaging, reflective, and dynamic conversational style compared to the nonrelational chatbot.

By contrast, participants in the nonrelational chatbot condition used more words indicating negation and agreement than those interacting with the relational chatbot. Considering their overall word usage was lower than in the relational chatbot condition, this suggests a generally more negative or resistant conversational tone, while simultaneously showing a tendency to agree with the chatbot without further elaboration, reflecting a terse conversational style.

In regard to establishing human–chatbot relationships, a marginal difference was observed in the "bond" dimension of the therapeutic alliance between the chatbot and participants, with individuals interacting with the relational chatbot reporting a stronger social bond compared to those in the control chatbot condition. This strengthens the notion in healthcare contexts that chatbots hold great value as they can emulate physicians' relational communicative behaviors, which have shown increased human–chatbot relationships[23,34] and significant health outcomes[37] in previous studies.

Throughout the 7-day intervention period, participants using the relational chatbot increased by 2748 steps, while those in the control chatbot condition increased by 3458 steps a week. Assuming that it takes roughly 10 min to take 1000 steps,[9] this increase in steps translates to an additional 27 to 34 min of physical activity accumulated over the course of the week. Taking into account that an increase of 1000 steps per day has been associated with a reduced risk of all-cause mortality in adults,[59] the observed outcomes of this 1-week intervention can be considered promising.

However, these findings should be interpreted with caution. While both conditions showed an overall increase in steps from baseline to the intervention period, the linear mixed model indicated a decreasing trend in step counts

Table 2. Results of linear mixed model on daily step counts (n = 36).

Effect	Estimate	SE	df	t	P-value	CI [lower, upper]
Intercept	5583.93	676.95	80.31	8.249	<.001	[4257.1, 6910.7]
Timepoint[a]	−243.98	118.47	248.33	−2.059	.041	[−476.2, −11.8]
Condition[b]	−1551.47	886.00	80.22	−1.751	.084	[−3288.0, 185.1]
Timepoint × Condition	301.08	154.42	247.91	1.950	.052	[−1.6, 603.7]

Note. [a]Baseline steps were collected for 7 days prior to intervention and included in the model as timepoint = 0; [b]Relational chatbot condition was coded as 1.
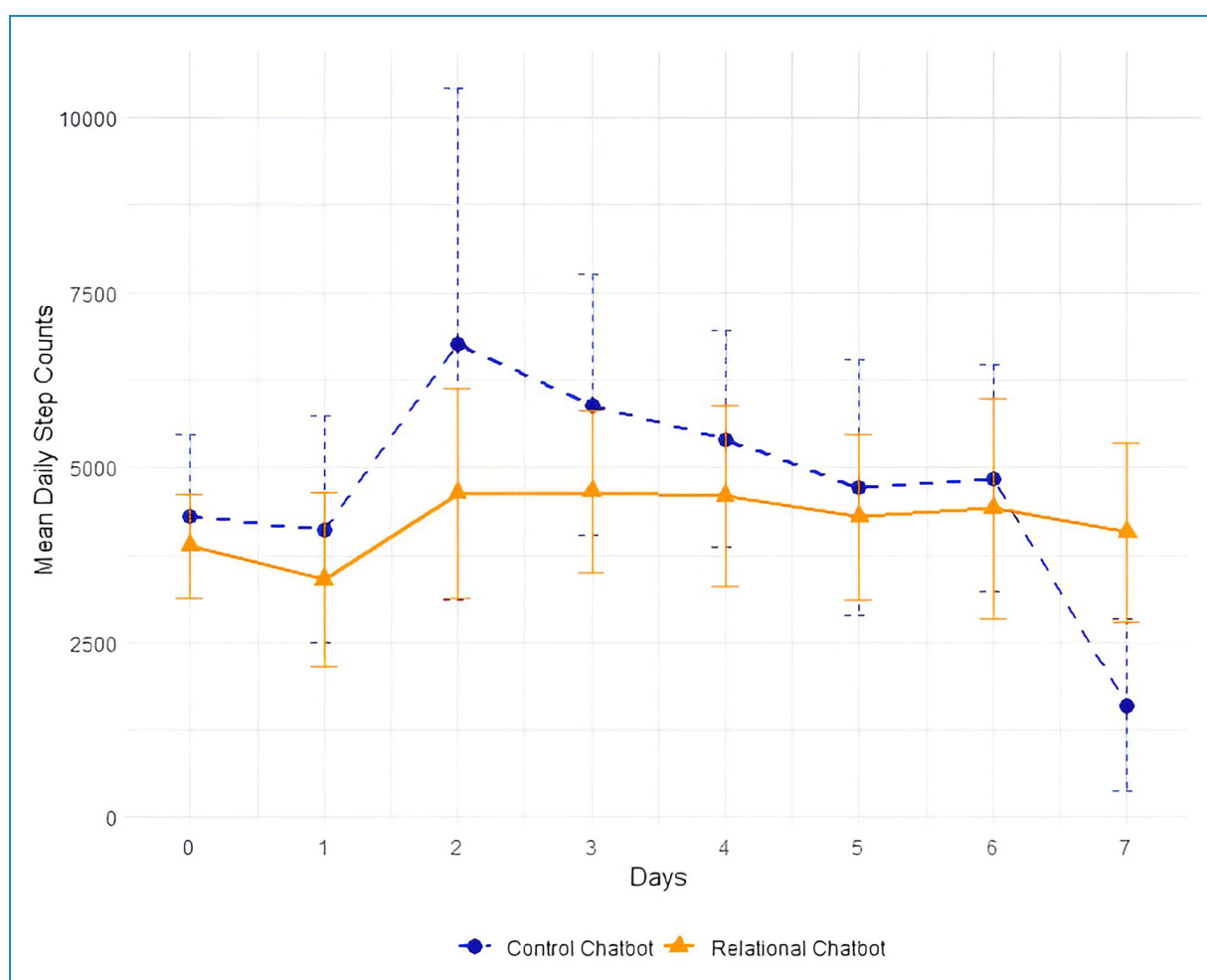
Figure 3. Mean daily step count and 95% confidence intervals by days of the study.

over time when accounting for the time factor. As shown in Figure 3, during the 7-day intervention, participants initially increased their activity in the first few days but showed a gradual decline. This may suggest that the initial increase in physical activity was driven by the novelty of interacting with an AI chatbot, and the similarity in improvement between groups suggests a potential digital placebo effect linked to user expectations of digital devices,[60] where the mere act of interacting with any chatbot, regardless of its relational capabilities, may have motivated increased activity.

A notable finding was that this decline in step counts was more pronounced in the nonrelational (control) condition, with a steep drop observed on the final day. In contrast,

participants in the relational chatbot condition maintain their step counts. This may suggest that the relational chatbot helped mitigate the postnovelty decline in physical activity levels and sustain participants' engagement.

In addition, although the nonrelational (control) group engaged more consistently over the days, a deeper analysis of the qualitative aspects of the participants' conversation and their word choices reveals important differences in processing mechanisms. The relational cues—social dialogue, self-disclosure, empathy, meta-relational, humor—may have triggered motivation by creating a more supportive and relatable environment. For example, participants interacting with the relational chatbot showed a greater use of various language features, which are indicative of a more thoughtful and reflective engagement style. Consequently, these sustained interactions may have fostered a positive reinforcement in maintaining physical activity over time.

In contrast, those interacting with a nonrelational (control) chatbot used more terse negations and assent words, indicative of a more surface-level interaction, suggesting that the participants may have approached the interaction with less depth or engagement. As such, while participants in nonrelational (control) condition show higher step counts over the six days, the pattern is less sustainable and ends in a decline in step counts on the final day. This may indicate that participants in the nonrelational (control) condition perceived physical activity as more of a "task," thereby losing interest as it neared completion. This contrast supports our interpretation that the relational chatbot may promote sustained engagement through relationship-building, rather than merely task completion.

Therefore, future research should consider these key areas for improvement. First, because novelty effects often diminish over time, longer-duration studies are needed to determine if the observed effects persist beyond the initial novelty period.[61] Second, to clearly isolate the impact of the chatbot's relational features, studies should consider incorporating an additional comparator, such as standard printed materials or no intervention[62] to provide a clearer baseline for assessing the true impact of AI chatbot interventions on sustained physical activity levels. Lastly, it would be informative to triangulate multiple indicators of sustained engagement. Beyond assessing physical activity levels, it would be valuable to analyze the quantity and quality of exchanged messages, the depth of participants' responses, as well as their emotional tone could provide a more comprehensive understanding of participants' engagement and interest over time.

Strengths and limitations

This study employed both objective and subjective measures of physical activity, thereby facilitating a comprehensive and accurate evaluation of the intervention's effectiveness.[63] Additionally, drawing upon existing literature on human–chatbot interactions in healthcare settings, this study introduced an AI chatbot capable of exhibiting more humanlike, relationship-building behaviors to enhance health outcomes. A strength of this study lies in its focus on the importance of relational components in AI chatbot-based health interventions. Unlike much of the existing research that primarily emphasizes quantitative outcomes, this study provides qualitative insights into how relational chatbots influence user engagement and conversational depth. The findings highlight that relational components of the chatbots are not merely supplementary but integral to fostering a sustained health engagement process and adherence. Nonetheless, this study presents certain limitations.

First, the sample was relatively small (n = 36) and consisted predominantly of college students, who generally represent a younger and more active population. This homogeneity in our sample may limit our ability to draw conclusions about the chatbot's usability and potential effectiveness across diverse age groups, activity levels, and socioeconomic backgrounds. The focus on college students may have introduced bias in regard to technology literacy, daily schedules, and motivations for physical activity, which may not be representative of the general population. For future research, it would be beneficial to expand the study to include a more diverse sample. Particularly, this should encompass participants with varying levels of digital literacy, ranging from those experienced with AI chatbots to individuals with low technological proficiency or even technology-resistant attitudes. Such an approach would significantly enhance the study's external validity, providing insights into the chatbot's acceptance and usability across a broader range of potential users.

In addition, the small sample size and relatively short intervention period (7 days) limited our ability to detect significant differences between intervention conditions and further examine the role of human–chatbot relationships on physical activity outcomes. To address this, future studies should employ a larger-scale design with an extended intervention period. Increasing the sample size would enhance statistical power, enabling detection of meaningful differences between conditions and allowing for mediation analyses. For instance, extending the intervention to 3 months with an additional 3-month follow-up period would allow for observation of both immediate effects and maintenance[64] of physical activity levels. These changes would provide more robust evidence of the chatbot's effectiveness and insights into the sustainability of AI-driven physical activity interventions.

Lastly, while pedometer data provide objective estimates of step counts, they do not capture all forms of daily physical activity. Future research should incorporate a more comprehensive approach to measuring physical activity, combining self-report questionnaires and/or diaries with

devices such as accelerometers and heart rate monitors.[65] This multi-method approach would provide a more complete picture of participants' activity levels and patterns than relying on a single measure.

Ethical considerations

Chatbots offer significant potential in healthcare and behavior change by providing support to vulnerable populations and tackling challenges related to the high costs of healthcare and behavioral therapy. Yet, they can also bring about new challenges and potential long-term consequences for individuals engaging with these services in healthcare and behavior change contexts. For instance, when a chatbot mimics a human therapist, individuals may develop certain expectations, potentially leading to dependency or blind trust, and anticipating more assistance than the chatbot is capable of providing.[66] Establishing relationships with artificial agents can lead to social and health benefits.[67] However, previous research has documented cases of emotional dependency that may result in addiction and negatively impact individuals' real-life intimate relationships.[68–70] Therefore, researchers need to find a balance in these interactions to minimize potential adverse effects arising from excessive attachment or blind trust.

Conclusions

The rapid advancement of AI chatbot technologies, particularly in relationship-building capacities, offers potentially promising applications in complementing human efforts to promote physical activity. While our findings are suggestive rather than definitive, the present study provides initial evidence that chatbots equipped with relational capabilities may have the potential to effectively motivate individuals to engage in physical activity. However, it should be noted that AI chatbots are not intended to replace human healthcare providers, but rather to complement and extend their efforts, potentially increasing the reach and frequency of interventions. Future research employing larger samples could yield more robust statistical power and further insights into user outcomes. Additionally, it would be beneficial for future studies to incorporate a control condition utilizing an alternative medium (e.g., a website) to establish a more natural health information-providing setting, facilitating a comparison of the effectiveness of chatbots versus other methods in promoting physical activity.

ORCID iD: Yoo Jung Oh https://orcid.org/0000-0002-7829-8535
Yoshimi Fukuoka https://orcid.org/0000-0002-2245-9264
Jingwen Zhang https://orcid.org/0000-0003-1733-6857

References

1. Kohler S, Dippon L, Helsper N, et al. Population-based physical activity promotion with a focus on health equity: a review of reviews. *Int J Equity Health* 2023 Jan 26; 22: 18.
2. Carlson SA, Fulton JE, Pratt M, et al. Inadequate physical activity and health care expenditures in the United States. *Prog Cardiovasc Dis* 2015 Jan; 57: 315–323.
3. U.S Department of Health and Human Services. *Physical Activity Guidelines for Americans*. 2nd ed. Washington, DC: U.S. Department of Health and Human Services, 2018.
4. Heath GW, Parra DC, Sarmiento OL, et al. Evidence-based intervention in physical activity: lessons from around the world. *Lancet* 2012 Jul; 380: 272–281.
5. Pereira J and Díaz Ó. Using health chatbots for behavior change: a mapping study. *J Med Syst* 2019 May; 43: 135.
6. Kowatsch T, Volland D, Shih I, et al. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. In: *Designing the Digital Transformation: 12th International Conference*, DESRIST 2017, Karlsruhe, Germany, May 30–June 1, 2017, Proceedings 12 2017, pp.485–489: Springer International Publishing.
7. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc* 2018 Sep 1; 25: 1248–1258.
8. Zhang J, Oh YJ, Lange P, et al. Artificial intelligence chatbot behavior change model for designing artificial intelligence chatbots to promote physical activity and a healthy diet: viewpoint. *J Med Internet Res* 2020 Sep 30; 22: e22845.
9. Maher CA, Davis CR, Curtis RG, et al. A physical activity and diet program delivered by artificially intelligent virtual health coach: proof-of-concept study. *JMIR mHealth uHealth* 2020 Jul 10; 8: e17558.
10. Amato F, Marrone S, Moscato V, et al. Chatbots meet eHealth: Automatizing healthcare. In *WAIAH@ AI* IA* 2017 Nov 14 (pp.40–49).

11. Sweeney C, Potts C, Ennis E, et al. Can chatbots help support a person's mental health? Perceptions and views from mental healthcare professionals and experts. *ACM Trans Comput Healthc* 2021 Jul 31; 2: 1–15.

12. Aggarwal A, Tam CC, Wu D, et al. Artificial intelligence–based chatbots for promoting health behavioral changes. Systematic review. *J Med Internet Res* 2023 Feb 24; 25: e40789.

13. Oh YJ, Zhang J, Fang ML, et al. A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. *Int J Behav Nutr Phys Act* 2021 Dec 11; 18: 160.

14. To QG, Green C and Vandelanotte C. Feasibility, usability, and effectiveness of a machine learning–based physical activity chatbot: quasi-experimental study. *JMIR mHealth uHealth* 2021 Nov 26; 9: e28577.

15. Bickmore TW and Picard RW. Establishing and maintaining long-term human-computer relationships. *ACM Trans Comput-Hum Interact* 2005 Jun; 12: 293–327.

16. Chandra S, Mohammadnezhad M and Ward P. Trust and communication in a doctor- patient relationship: a literature review. *Arch Med* 2018; 3: 36.

17. Han R, Todd A, Wardak S, et al. Feasibility and acceptability of chatbots for nutrition and physical activity health promotion among adolescents: systematic scoping review with adolescent consultation. *JMIR Hum Factors* 2023 May 5; 10: e43227.

18. Duarte-Diaz A, Perestelo-Perez L, Gelabert E, et al. Efficacy, safety, and evaluation criteria of mHealth interventions for depression: systematic review. *JMIR Mental Health* 2023 Sep 27; 10: e46877.

19. Waller R and Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. *Psychol Med* 2009 May; 39: 705–712.

20. Hoermann S, McCabe KL, Milne DN, et al. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. *J Medical Internet Res* 2017 Jul 21; 19: e267.

21. Schachner T, Keller R and Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. *J Medical Internet Res* 2020 Sep 14; 22: e20701.

22. Brave S, Nass C and Hutchinson K. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. *Int J Hum-Comput Stud* 2005 Feb 1; 62: 161–178.

23. Kraus M, Seldschopf P and Minker W. Towards the development of a trustworthy chatbot for mental health applications. In: *Multi Media Modeling: 27th International Conference*, MMM 2021, Prague, Czech Republic, June 22–24, 2021, Proceedings, Part II 27 2021, pp.354–366: Springer International Publishing.

24. Mattar N and Wachsmuth I. Small talk is more than chit-chat: exploiting structures of casual conversations for a virtual agent. In: *Annual Conference on Artificial Intelligence 2012* Sep 24, pp.119–130. Berlin, Heidelberg: Springer Berlin Heidelberg.

25. Rhim J, Kwak M, Gong Y, et al. Application of humanization to survey chatbots: change in chatbot perception, interaction experience, and survey data quality. *Comput Hum Behav* 2022 Jan; 126: 107034.

26. Gelso CJ and Hayes JA. *The psychotherapy relationship: Theory, research, and practice*. New York, NY: John Wiley & Sons Inc, 1998.

27. Fuertes JN, Mislowack A, Bennett J, et al. The physician–patient working alliance. *Patient Educ Couns* 2007 Apr; 66: 29–36.

28. Graffigna G, Barello S and Bonanomi A. The role of patient health engagement model (PHE-model) in affecting patient activation and medication adherence: a structural equation model. *PloS One* 2017 Jun 27; 12: e0179865.

29. Wright JH and Davis D. The therapeutic relationship in cognitive-behavioral therapy: patient perceptions and therapist responses. *Cognit Behav Pract* 1994 Jun 1; 1: 25–45.

30. Elvins R and Green J. The conceptualization and measurement of therapeutic alliance: an empirical review. *Clin Psychol Rev* 2008 Oct; 28: 1167–1187.

31. Gelso CJ and Carter JA. The relationship in counseling and psychotherapy: components, consequences, and theoretical antecedents. *Couns Psychol* 1985 Apr 1; 13: 155–243.

32. Weiss KA, Smith TE, Hull JW, et al. Predictors of risk of non-adherence in outpatients with schizophrenia and other psychotic disorders. *Schizophr Bull* 2002 Jan 1; 28: 341–349.

33. Lee YC, Yamashita N, Huang Y, et al. '"I hear you, I feel you': encouraging deep self-disclosure through a chatbot. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 2020* Apr 21, pp.1–12.

34. Bickmore TW, Mitchell SE, Jack BW, et al. Response to a relational agent by hospital patients with depressive symptoms. *Interact Comput* 2010 Jul; 22: 289–298.

35. Salman S, Richards D and Caldwell P. Analysis of empathic dialogue in actual doctor-patient calls and implications for design of embodied conversational agents. *Ital J Comput Linguist* 2021 Dec 1; 7: 91–112.

36. Bickmore TW, Caruso L, Clough-Gorr K, et al. It's just like you talk to a friend' relational agents for older adults. *Interact Comput* 2005 Dec; 17: 711–735.

37. Bickmore TW, Schulman D and Sidner C. Automated interventions for multiple health behaviors using conversational agents. *Patient Educ Couns* 2013 Aug; 92: 142–148.

38. Roller S, Dinan E, Goyal N, et al. Recipes for building an open-domain chatbot. arXiv preprint arXiv:2004.13637. 2020 Apr 28.

39. Fukuoka Y, Haskell W, Lin F, et al. Short- and long-term effects of a mobile phone app in conjunction with brief in-person counseling on physical activity among physically inactive women: the mPED randomized clinical trial. *JAMA Netw Open* 2019 May 24; 2: e194281.

40. Fukuoka Y, Komatsu J, Suarez L, et al. The mPED randomized controlled clinical trial: applying mobile persuasive technologies to increase physical activity in sedentary women protocol. *BMC Pub Health* 2011 Dec; 11: 933.

41. Fukuoka Y, Haskell W and Vittinghoff E. Mechanisms of an app-based physical activity intervention and maintenance in community-dwelling women: mediation analyses of a randomized controlled trial. *J Cardiovasc Nurs* 2023 Mar 1; 38: E61–E69.

42. Liang KH, Lange P, Oh YJ, et al. Evaluation of in-person counseling strategies to develop physical activity chatbot for women. arXiv preprint arXiv:2107.10410. 2021 Jul 22.

43. Bandura A, Adams NE and Beyer J. Cognitive processes mediating behavioral change. *J Pers Soc Psychol* 1977; 35: 125–139.

44. Tudor-Locke C, Hatano Y, Pangrazi RP, et al. Revisiting "how many steps are enough?" *Med Sci Sports Exerc* 2008 Jul; 40: S537–S543.

45. Liao Y and He J. Racial mirroring effects on human-agent interaction in psychotherapeutic conversations. In: *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 2020 Mar 17, pp.430–442.

46. Griol D, Molina JM and Sanchis A. A multimodal conversational coach for active ageing based on sentient computing and m-health. *Expert Syst* 2020 Apr; 37: e12454.

47. Holmes S, Moorhead A, Bond R, et al. 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 Sep 10, pp.207–214.

48. Kocielnik R, Xiao L, Avrahami D, et al. Reflection companion: a conversational system for engaging users in reflection on physical activity. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2018 Jul 5; 2: 1–26.

49. Pennebaker JW, Boyd RL, Jordan K, et al. The development and psychometric properties of LIWC2015.

50. Tausczik YR and Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. *J Lang Soc Psychol* 2010 Mar; 29: 24–54.

51. McEwen WJ and Greenberg BS. The effects of message intensity on receiver evaluations of source, message and topic. *J Commun* 1970 Dec; 20: 340–350.

52. Hatcher RL and Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. *Psychother Res* 2006 Jan; 16: 12–25.

53. Horvath AO and Greenberg LS. Development and validation of the working alliance inventory. *J Couns Psychol* 1989 Apr; 36: 223.

54. Craig C, Marshall A, Sjostrom M, et al. International physical activity questionnaire-short form. *J Am Coll Health* 2017; 65: 492–501.

55. Kramer JN, Künzler F, Mishra V, et al. Which components of a smartphone walking app help users to reach personalized step goals? Results from an optimization trial. *Ann Behav Med* 2020 Jun 12; 54: 518–528.

56. Rhodes RE, Courneya KS and Hayduk LA. Does personality moderate the theory of planned behavior in the exercise domain? *J Sport Exerc Psychol* 2002 Jun; 24: 120–132.

57. Piao M, Ryu H, Lee H, et al. Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: exploratory randomized controlled trial. *JMIR mhealth uhealth* 2020 May 19; 8: e15085.

58. Hardeman W, Houghton J, Lane K, et al. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *Int J Behav Nutr Phys Act* 2019; 16: 1–21.

59. Hall KS, Hyde ET, Bassett DR, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. *Int J Behav Nutr Phys Act* 2020 Dec; 17: 78.

60. Torous J and Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. *Lancet Psychiatry* 2016 Feb; 3: 100–102.

61. Croes EA and Antheunis ML. Can we be friends with Mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *J Soc Pers Relat* 2021 Jan; 38: 279–300.

62. Direito A, Carraça E, Rawstorn J, et al. Mhealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. *Ann Behav Med* 2017 Apr 1; 51: 226–239.

63. Skender S, Ose J, Chang-Claude J, et al. Accelerometry and physical activity questionnaires - a systematic review. *BMC Pub Health* 2016 Jun 16; 16: 515.

64. Fjeldsoe B, Neuhaus M, Winkler E, et al. Systematic review of maintenance of behavior change following physical activity and dietary interventions. *Health Psychol* 2011 Jan; 30: 99.

65. Sylvia LG, Bernstein EE, Hubbard JL, et al. A practical guide to measuring physical activity. *J Acad Nutr Diet* 2014 Feb; 114: 199.

66. Sedlakova J and Trachsel M. Conversational artificial intelligence in psychotherapy: a new therapeutic tool or agent? *Am J Bioeth* 2023 May 4; 23: 4–13.

67. Prochaska JJ, Vogel EA, Chieng A, et al. A therapeutic relational agent for reducing problematic substance use (woebot): development and usability study. *J Med Internet Res* 2021 Mar 23; 23: e24850.

68. Pentina I, Hancock T and Xie T. Exploring relationship development with social chatbots: a mixed-method study of replika. *Comput Hum Behav* 2023 Mar; 140: 107600.

69. Laestadius L, Bishop A, Gonzalez M, et al. Too human and not human enough: a grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. *New Media Soc* 2022 Dec 22; 26: 5923–5941.

70. Xie T and Pentina I. Attachment theory as a framework to understand relationships with social Chatbots: a case study of Replika. In: *Proceedings of the 55th Hawaii International Conference on System Sciences*, 2022.