Enhancing Physical Activity Among University Students in China: A Parallel-Group Quasi-Randomized Controlled Trial of an AI Chatbot Intervention Assessing Engagement, Usability, and Efficacy

Bing Xu, Chao Liu, Han Liu, Hai Li, Yong-Yee Chong, Ling Qin, Zheng Xu, Ke Ma, Andre Matthias Müller

Submitted to: Journal of Medical Internet Research
on: August 31, 2025

Table of Contents

Enhancing Physical Activity Among University Students in China: A Parallel-Group Quasi-Randomized Controlled Trial of an AI Chatbot Intervention Assessing Engagement, Usability, and Efficacy

Bing Xu[1, 2]; Chao Liu[1]; Han Liu[1]; Hai Li[3]; Yong-Yee Chong[2]; Ling Qin[2, 4]; Zheng Xu[1]; Ke Ma[1]; Andre Matthias Müller[5]

[1] School of Physical Education Jiangxi University of Technology Nanchang CN
[2] Faculty of Sports and Exercise Science Universiti Malaya Kuala Lumpur MY
[3]Teaching and Research Section of Physical Education College of Sport Neijiang Normal University NeiJiang CN
[4] Faculty of Physical Education and Health Science Chongqing Normal University ChongQing CN
[5] Saw Swee Hock School of Public Health National University of Singapore Singapore SG

Corresponding Author:
Bing Xu

School of Physical Education
Jiangxi University of Technology
ZiYangDaDao 115
Nanchang
CN

Abstract

Background: Physical inactivity among university students is a prevalent global health concern with long-term implications. Emerging technologies, such as wearable devices and AI-driven chatbots, offer promising avenues to promote physical activity (PA). However, empirical studies assessing the engagement, usability, and efficacy of next-generation large language model-based AI chatbots remain limited.

Objective: This study aimed to evaluate the engagement, usability, and efficacy of an AI Coach intervention, based on the Doubao large language model, in enhancing PA levels among university students in China.

Methods: A parallel-group quasi-randomized controlled trial was conducted over 8 weeks with 60 eligible university student participants (intervention group, n=30; control group, n=30). The intervention group used both their wearable devices and the AI Coach delivered via WeChat, while the control group used wearable devices only. Outcome measures included system-recorded interaction metrics to assess engagement, the System Usability Scale (SUS) and user experience questionnaires for usability, and changes in average daily step counts and weekly moderate-to-vigorous physical activity (MVPA) duration for efficacy. Data were analyzed using mixed-design ANOVA and appropriate non-parametric tests.

Results: Participants engaged actively with the AI Coach, exchanging an average of 17.20 conversation rounds (SD = 8.72), with 60.27% of conversations initiated by participants. The average conversation duration was approximately 3 minutes. The mean SUS score was 85.75 (SD = 3.99), indicating excellent usability. Compared to the control group, the intervention group showed significant increases in average daily steps (mean increase 557 steps; $p < 0.001$) and weekly MVPA duration (mean increase 8.73 minutes; $p < 0.001$). No adverse events were reported.

Conclusions: The Doubao-based AI Coach is a safe, practical, and efficacious tool to enhance PA among university students. Its high engagement and usability suggest potential for scalable digital health interventions. This study highlights the value of combining AI chatbots with wearable devices to deliver personalized, real-time coaching. Future research should focus on automating data collection, expanding to diverse populations, and exploring user experiences to improve the intervention's effectiveness and broader public health impact. Clinical Trial: Ethical approval was obtained from the Research Ethics Committee of Jiangxi University of Technology (JXUT.TNC2/JXUTREC2409).

(JMIR Preprints 31/08/2025:83307)
DOI: https://doi.org/10.2196/preprints.83307

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

- ✓ **Please make my preprint PDF available to anyone at any time (recommended).**
- Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
- Only make the preprint title and abstract visible.
- No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

- ✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
- Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
- Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http
- No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <a href="https:/

Original Manuscript

Enhancing Physical Activity Among University Students in China: A Parallel-Group Quasi-Randomized Controlled Trial of an AI Chatbot Intervention Assessing Engagement, Usability, and Efficacy

Abstract

Background: Physical inactivity among university students is a prevalent global health concern with long-term implications. Emerging technologies, such as wearable devices and AI-driven chatbots, offer promising avenues to promote physical activity (PA). However, empirical studies assessing the engagement, usability, and efficacy of next-generation large language model-based AI chatbots remain limited.

Objective: This study aimed to evaluate the engagement, usability, and efficacy of an AI Coach intervention, based on the Doubao large language model, in enhancing PA levels among university students in China.

Methods: A parallel-group quasi-randomized controlled trial was conducted over 8 weeks with 60 eligible university student participants (intervention group, n=30; control group, n=30). The intervention group used both their wearable devices and the AI Coach delivered via WeChat, while the control group used wearable devices only. Outcome measures included system-recorded interaction metrics to assess engagement, the System Usability Scale (SUS) and user experience questionnaires for usability, and changes in average daily step counts and weekly moderate-to-vigorous physical activity (MVPA) duration for efficacy. Data were analyzed using mixed-design ANOVA and appropriate non-parametric tests.

Results: Participants engaged actively with the AI Coach, exchanging an average of 17.20 conversation rounds (SD = 8.72), with 60.27% of conversations initiated by participants. The average conversation duration was approximately 3 minutes. The mean SUS score was 85.75 (SD = 3.99), indicating excellent usability. Compared to the control group, the intervention group showed significant increases in average daily steps (mean increase 557 steps; $p < 0.001$) and weekly MVPA duration (mean increase 8.73 minutes; $p < 0.001$). No adverse events were reported.

Conclusions: The Doubao-based AI Coach is a safe, practical, and efficacious tool to enhance PA among university students. Its high engagement and usability suggest potential for scalable digital health interventions. This study highlights the value of combining AI chatbots with wearable devices to deliver personalized, real-time coaching. Future research should focus on automating data collection, expanding to diverse populations, and exploring user experiences to improve the intervention's effectiveness and broader public health impact.

Keywords

AI Chatbot; Physical Activity; University Students; Engagement; Usability; Efficacy

Introduction

Regular physical activity (PA) is crucial for maintaining overall health, offering benefits such as strengthening the immune system, reducing the risk of non-communicable diseases, and improving mental well-being. According to the World Health Organization's (WHO) Guidelines on Physical Activity and Sedentary behavior, adults are advised to engage in 150-300 minutes of moderate-intensity aerobic activity, 75-150 minutes of vigorous-intensity activity, or an equivalent combination per week [1]. Daily step count is a readily accessible and practical metric for monitoring and setting PA goals [2]. Paluch et al. demonstrated that mortality risk progressively decreases with increasing daily steps, with the risk reduction plateauing at approximately 8,000–10,000 steps among younger adults[3]. Conversely, physical inactivity and prolonged sedentary behavior have been associated with various adverse health outcomes, including obesity, cancer, diabetes, and depression [4–7]. Data derived from 358 studies across 168 countries also revealed that 27.5% of 1.9 million participants did not meet the recommended PA levels [8]. Given its widespread prevalence and health implications, physical inactivity represents an urgent global health issue that necessitates targeted intervention research.

University students represent a key population for PA interventions, as multiple studies worldwide have consistently reported insufficient PA within this group. For example, 65% of Canadian university students do not meet the recommended PA guidelines for health [9]. Similarly, approximately 41.4% of undergraduate students in Malaysia are physically inactive [10]. In China, 24.4% of university students are classified as physically inactive, while 18.2% of university students in Turkey are reported as inactive [11,12]. Furthermore, one in three university students in Romania have been found to be inactive [13]. These findings underscore a global pattern of low PA levels within university populations.

Physical inactivity during young adulthood has serious consequences for long-term health, contributing to increased risks of chronic diseases and reduced quality of life. Lifestyle habits established during the university years tend to persist into adulthood and have significant implications for long-term health. The health-related behaviors formed during this transitional period can influence risk factors for chronic conditions and overall well-being later in life [14]. Additionally, university students represent the upcoming generation of leaders, decision-makers, and parents; therefore, fostering healthy lifestyle habits at this stage contributes to shaping future societal health outcomes [15]. Establishing a foundation of healthy behaviors during university is critical for promoting optimal middle-age health, minimizing the prevalence of suboptimal health conditions, and potentially delaying the onset of chronic diseases [16].

Factors influencing university students' PA include individual factors (e.g., self-efficacy, perceived barriers), interpersonal support (e.g., from parents, peers, and teachers), organizational aspects (e.g., school facilities and policies), community-level determinants (e.g., built environment and access to services), and national policies that influence behavior both directly and indirectly [17]. Current strategies to enhance PA levels among university students typically involve health education initiatives, increased utilization of campus sports facilities, the promotion of healthy lifestyle behaviors, and the establishment of supportive social networks. Although these interventions have demonstrated some promise, overall PA levels in this population remain suboptimal [18].

Despite efforts targeting individual, social, and environmental factors, improving physical activity levels among university students remains difficult. Traditional interventions show limited success in overcoming complex barriers. As a tech-savvy group, university students are receptive to emerging digital technologies like wearable devices and artificial intelligence (AI), which can provide personalized, real-time support to promote healthier behaviors. University students generally show high acceptance of emerging technologies such as wearable devices and AI [19,20]. This enthusiasm aligns with growing evidence that these tools have significant potential to promote and enhance physical activity. These technologies can provide personalized feedback and real-time guidance, which enhance user engagement and facilitate sustained behavior change. A recent meta-analysis reported that the use of wearable devices resulted in an average increase of 1,312.23 steps per day and an additional 57.8 minutes of moderate-to-vigorous physical activity (MVPA) per week [21]. Evidence also suggests that AI-driven chatbots hold substantial promise for promoting healthy lifestyles and increasing PA, as they can tailor interventions by interpreting users' contexts, engaging in persuasive dialogues, and delivering feedback based on behavioral data [21,22].

AI chatbots are conversational agents that leverage AI and machine learning technologies to simulate human interactions by engaging users in natural language dialogues [23]. As tools for health behavior intervention, AI chatbots demonstrate feasibility and initial efficacy in guiding users to modify lifestyle habits [22,24,25]. For instance, Kocielnik and Hsieh employed a rule-based chatbot to prompt users to reflect on their PA data [26]. Similarly, chatbot applications such as Ally have been employed to assist participants in achieving personalized step goals [27,28]. Piao et al. [29] developed a chatbot-based program designed to encourage office workers to adopt the habit of using stairs, while Maher et al. [24] implemented an AI-driven virtual health coach that integrated Mediterranean diet and PA interventions to improve both PA levels and dietary habits. Furthermore, Tess, an AI chatbot, has been successfully employed in supporting the management and treatment of adolescent obesity [30]. Recent studies, including Singh et al. [31], have further examined the efficacy of AI chatbots in increasing both PA and healthy eating behaviors.

The launch of ChatGPT 3.5 in 2022 marked the beginning of a new generation of chatbots. Featuring enhanced natural language processing, context-aware reasoning, and self-explanatory outputs that surpass previous models in conversational depth and task adaptability, ChatGPT 3.5 delivers more intelligent and personalized interactive experiences, positioning it as one of the fastest-growing consumer applications [32–35]. Subsequently, similar products, including Gemini, Grok, Doubao, and Kimi, have been introduced, thereby expanding access to this technology more cost-effectively and conveniently [36–39].

Despite these advancements, there remains a notable gap in empirical studies that utilize new generation AI chatbots based on large language models to enhance PA. Much of the previous research has combined chatbots with fitness trackers as a single intervention, with only one study comparing the use of a fitness tracker (Fitbit) alone versus a combination of a fitness tracker and a chatbot [40]. To address this gap, the present study aims to enhance PA levels among university students in China and evaluate the engagement, usability, and efficacy of an AI Coach intervention based on these next-generation chatbots.

Materials and methods

This study was a parallel-group quasi-randomized controlled trial conducted from August to October 2024 over 8 weeks. Eligible participants were consecutively enrolled and assigned to intervention or control groups based on odd/even study identification numbers assigned upon screening completion. The intervention group received the AI Coach alongside their wearable device, whereas the control group used only their wearable device. Participants were blinded to group allocation and informed only that their physical activity data would be collected. Ethical approval was obtained from the Research Ethics Committee of Jiangxi University of Technology (JXUT.TNC2/JXUTREC2409), and informed consent was obtained from all participants.

Participants

Participants were recruited between August and September 2024 through social media platforms, in-person outreach, and snowball sampling techniques. Inclusion criteria were as follows: participants had worn a PA tracking device (e.g., smart bands or watches from Xiaomi, Huawei, Apple, Garmin, etc.) for at least 50 days within the preceding 60 days (days with fewer than 1,000 steps were considered invalid and excluded from analysis), were willing to continue wearing the device everyday throughout the study, were able to transmit PA data recorded by their device to a smartphone connected to the internet and share it with researchers via an app (e.g., Huawei Health), had a baseline daily step count of less than 8,000 steps and a weekly MVPA duration of less than 150 minutes, as measured by their device in the 7 days before the intervention, were able to use WeChat as an instant messaging software to send and receive text messages with others, and were at least 18 years of age. Exclusion criteria were as follows: participants who had medical or psychological conditions that contraindicated physical activity, or who were currently participating in other physical activity-related research projects.

Participants first completed an online screening survey, followed by a brief phone interview to confirm eligibility, address any potential questions, and schedule their baseline appointment. Before commencing the study, all participants reviewed the information sheet and were allowed to ask questions before providing informed consent. Eligible participants were required to sign a consent form and complete an online form, providing their age, gender, height, weight, and PA data (e.g., steps and MVPA duration) from the previous 7 days, as recorded through the corresponding apps (e.g., Huawei Health).

Participants were consecutively enrolled and assigned identification numbers in the order they completed screening. Participants with odd numbers were allocated to the intervention group, and those with even numbers were allocated to the control group. Participants were blinded to their group allocation and only informed that their physical activity data would be recorded.

Intervention

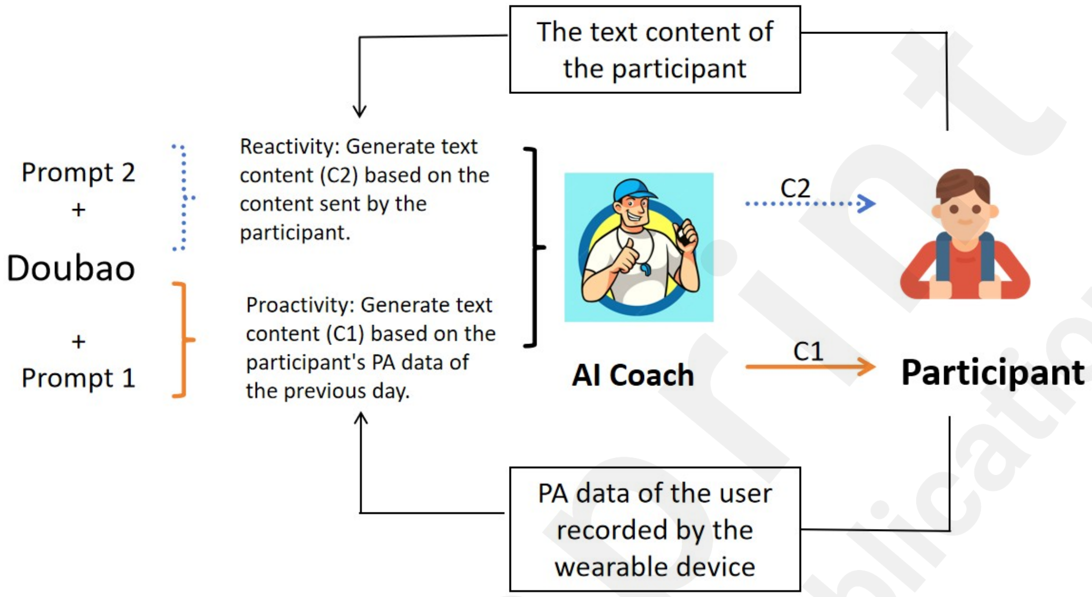
Participants in the intervention group used both their wearable device and the AI Coach, while those in the control group used only their wearable device. The intervention group began using the AI Coach by adding the Coach's WeChat account, which generates content via Doubao.

The AI chatbot, referred to as AI Coach, was developed using the large language model Doubao (version Doubao-pro-4k-240515, released in May 2024) [41]. It was integrated into the WeChat platform through the use of Java technology. The WeChat platform was selected due to its widespread popularity in China, ensuring that participants would not need to download a separate application to access the AI Coach. Participants' wearable devices (including brands such as Huawei, Xiaomi, OPPO, Apple, Garmin, etc.) were utilized to monitor their step counts and minutes in MVPA. The validity of step count measurements from these devices is well established; however, recent studies indicate that although MVPA estimation by consumer-grade devices has improved, it still exhibits systematic bias and should be interpreted with caution [42–44]. The step count and MVPA data collected by these devices were shared with the researchers via the respective brand applications (e.g., Huawei Health) and subsequently transmitted to the AI Coach.

As a behavioral intervention tool, the AI Coach was designed based on the COM-B model (capability, opportunity, motivation, and behavior) from the behavior Change Wheel [45]. The messages delivered by the AI Coach are designed to enhance participants' motivation (e.g., preemptively experiencing the satisfaction of completing a run), capability (e.g., understanding the benefits of reaching the daily 8,000-step goal), and opportunity (e.g., utilizing available facilities and equipment for exercise).

Figure 1 presents the proactive and reactive actions offered by the chatbot to support the participants.

Figure 1. Schematic diagram of AI coaching technology principles



Proactive actions involve the AI Coach generating content (C1) based on participants' PA data and providing suggestions for achieving daily goals, typically sent before 9:00 AM after retrieving the previous day's data. Reactive actions occur when participants send inquiries to the chatbot. Upon receiving content (C1), participants may respond to the AI Coach or send messages at any time, prompting the AI Coach to generate and send content (C2), typically including information about the benefits of PA and answers to specific exercise-related questions (e.g., exercises to do in the dormitory on rainy days). The proactive and reactive actions are defined through two sets of prompts for the AI Coach, which processes the prompts each time content is generated, integrating the received content to generate the appropriate response.

The prompts used in this study adhered to the TRACE framework, a widely recognized approach for crafting prompts, ensuring that generative AI products, such as the AI Coach, produce high-quality content that aligns with users' expectations [46].

Table 1. Key contents of prompt words for proactive and reactive actions of AI coach under TRACE framework.

TRACE Framework	Prompt 1 (Proactive)	Prompt 2 (Reactive)
(T) Title of Job	Fitness Coach	Fitness Coach
(R) Request	Analyze the user's PA data from the previous day and encourage them to achieve their daily target of 8,000 steps and 150 minutes of MVPA weekly.	Encourage the user based on their feedback and provide professional advice to help them achieve their daily target of 8,000 steps and 150 minutes of MVPA weekly.

(A) Audience	University students who have averaged fewer than 8,000 steps per day and less than 150 minutes of MVPA per week in the past 30 days.	University students who have averaged fewer than 8,000 steps per day and less than 150 minutes of per week in the past 30 days.
(C) Constraints	Communicate in a friendly yet professional tone.	Communicate in a friendly yet professional tone, using the COM-B model to analyze the user’s input and provide professional, reliable information and advice.
(E) Examples	Hello! Your average daily step count in the last 7 days was 7,425, which is slightly below the 8,000–10,000 step goal, but you're getting close! Your weekly MVPA time is 167 minutes, which meets the weekly goal of 150 minutes—well done! Consider incorporating more walking into your daily routine, like taking the stairs instead of the elevator. Keep up the good work!	Rainy weather can indeed affect outdoor activities, but there are still alternatives. On rainy days, you can try some indoor exercises to increase your step count and MVPA duration. Here are some suggestions: you can follow a 10-minute fat-burning or muscle-strengthening workout guided by your fitness app, or simply march in place in an obstacle-free area to stay active. If you'd like more suggestions, feel free to let me know!

Outcomes

Data collected during the 8-week intervention—including system-recorded interaction metrics, daily step counts, weekly MVPA duration, and post-intervention survey responses—were used to evaluate the engagement, usability, and efficacy of the AI Coach. To ensure the smooth running of the study, research assistants provided support for participants with any technical issues that arose during the intervention period.

Engagement was defined as the extent to which the AI Coach could be successfully implemented and sustain participant engagement. It was assessed using the following interaction metrics:

- **Exchanges**: quantified by the total number of messages exchanged and the total number of conversation rounds (a round defined as a response sent within 30 minutes of receiving a message from the AI Coach);
- **Duration**: measured by the length of conversations between participants and the AI Coach, including shortest, longest, average, and median conversation durations;
- **Initiation**: assessed by counting the number of conversation rounds initiated by the AI Coach versus those initiated by participants.

These indicators collectively reflect whether the system effectively attracted and maintained user participation throughout the intervention.

Usability was defined as participants' perceptions of the AI Coach's ease of use, helpfulness, and overall satisfaction. It was evaluated through:

- A user experience questionnaire assessing perceived helpfulness and any adverse reactions;
- The System Usability Scale (SUS), a validated 10-item instrument providing a subjective measure of system usability. SUS scores range from 0 to 100, with higher scores indicating better usability. Scores above 68 are considered above average, and those above 80 indicate excellent usability [47].

Efficacy was evaluated by comparing participants' PA levels (average daily step count and MVPA duration over the past 7 days) before and after using the AI Coach. If the wearable device recorded fewer than 1,000 steps on any given day, the device was considered not fully worn, and the data for that day were considered invalid.

Data Analysis

Demographic Characterization

Continuous variables were described using the mean ± standard deviation (mean ± SD) along with their minimum and maximum ranges [*min–max*]. All continuous variables were assessed for normality using the Shapiro–Wilk test before between-group comparisons (CG vs. AI). For data conforming to a normal distribution, independent sample t-tests were employed; when the assumption of normality was violated, Mann–Whitney U tests were used for non-parametric comparisons. Sex was treated as a categorical variable and expressed as a frequency (n). Between-group comparisons were performed using Fisher's exact test.

Analysis of behavioral Performance Indicators

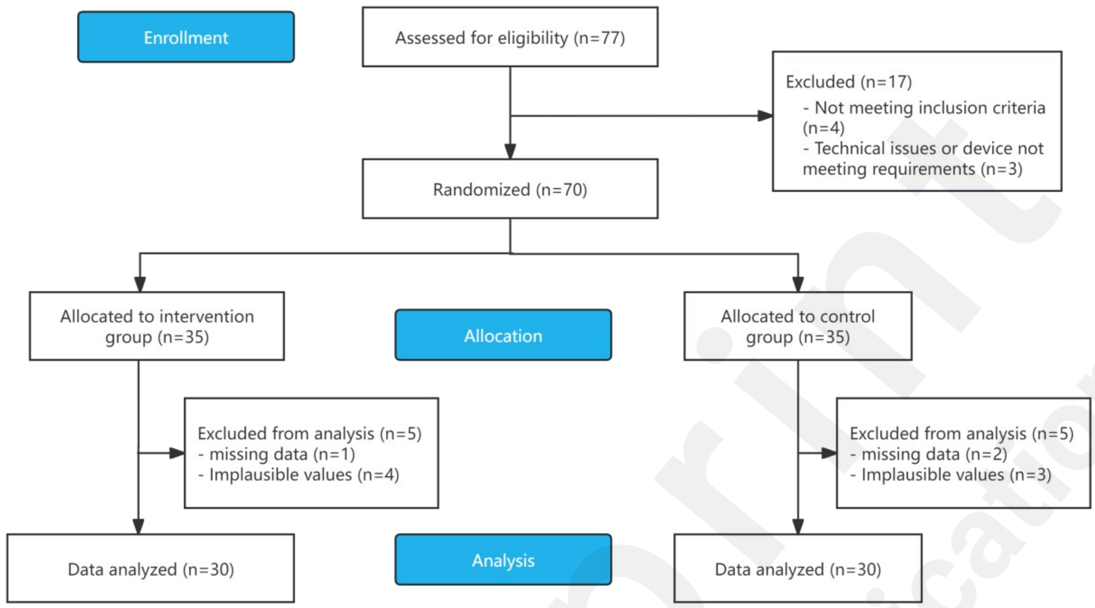
To assess the impact of AI chatbot interventions on PA, the study primarily used a 2 (group: IG vs. CG) × 2 (time: pre-intervention vs. post-intervention) mixed-design analysis of variance (ANOVA) to evaluate the behavioral performance indicators (including step counts and MVPA).
Before implementing the mixed-design ANOVA, the normality assumption for each variable within groups was tested using the Shapiro–Wilk test, and the homogeneity of variances was assessed using Levene's test to determine whether the assumptions required for ANOVA were satisfied.
If the data met the normality assumption but violated the homogeneity of variance assumption, Welch's ANOVA was used as an alternative to compare between groups, while within-group (pre–post) comparisons were performed using paired t-tests.
If the data violated the normality assumption, non-parametric methods were used: the Wilcoxon signed-rank test for within-group (pre–post) comparisons, and the Mann–Whitney U test for between-group comparisons.
All data analyses were performed using **R** software (version 4.4.2). Data cleaning was carried out using the **dplyr** package, and the mixed-design ANOVA was implemented using the ***ezANOVA*** function from the **ez** package. Normality testing was conducted using the ***shapiro.test*** function from the **stats** package, and variance homogeneity was assessed using the ***leveneTest*** function from the **car** package. Non-parametric tests, including the Mann–Whitney U test and Wilcoxon signed-rank test, were conducted using the ***wilcox.test()*** function from the **stats** package.

Results

Recruitment and Retention

Figure 2 illustrates the recruitment and retention process for the study participants.

Figure 2. Flow Diagram of Study Participants



A total of 77 participants were assessed for eligibility, of which 7 were excluded. The exclusion reasons included failure to meet inclusion criteria (n = 4) and technical issues or devices not meeting study requirements (n = 3). 70 participants who met eligibility criteria were enrolled consecutively. Each new enrollee was assigned a study identification number in order of screening completion; those with odd numbers were allocated to the intervention group (n = 35) and those with even numbers to the control group (n = 35). Participants remained unaware of study arm definitions and were only informed that their PA data would be recorded. All participants allocated to the intervention group received the assigned intervention and completed the study. Similarly, all participants in the control group continued using their wearable devices without intervention. For data analysis, ten participants (five from each group) were excluded due to either (1) missing data—defined as one or more days without any recorded step count or MVPA data (intervention group: n = 1; control group: n = 2), or (2) implausible values—defined as step count < 100 on one or more days (intervention group: n = 4; control group: n = 3). Therefore, data from 60 participants (intervention group: n = 30; control group: n = 30) were included in the final analysis.

Participant Characteristics

The demographic information of the participants is presented in Table 2.

Table 2. Baseline demographic characteristics of participants

Variable	CG (n = 30)	IG (n = 30)	p	Method

Mean age (years, SD) [range]	19.03 (0.76) [18.00–20.00]	18.97 (0.85) [18.00–20.00]	0.754	Mann-Whitney U test
Sex (Female / Male)	13 / 17	15 / 15	0.796	Fisher's exact test
Mean height (cm, SD) [range]	169.57 (6.67) [159.00–179.00]	170.17 (8.22) [158.00–188.00]	0.796	Mann-Whitney U test
Mean weight (kg, SD) [range]	58.73 (7.95) [47.00–72.00]	60.67 (8.73) [48.00–75.00]	0.359	Mann-Whitney U test
Mean BMI (kg/m², SD) [range]	20.31 (1.20) [18.13–22.47]	20.82 (1.09) [19.23–22.64]	0.078	Mann-Whitney U test

As shown in Table 2, the mean age was 19.03 years (SD = 0.76) in CG (n = 30) and 18.97 years (SD = 0.85) in IG (n = 30). The gender distribution was 13 females and 17 males in the CG and 15 females and 15 males in the IG, respectively. Differences in basic demographic variables, including sex, height, weight, and body mass index (BMI), were not statistically significant ($p > 0.05$), indicating comparable demographic characteristics between the two groups.

Engagement and Usability

Table 3 provides an overview of the interactions between the AI Coach and participants.

Table 3. Exchanges, duration, and initiation results

Variable	Total	Mean (SD)
Exchanges		
Total number of messages exchanged	3081	-
Total conversation (rounds)	516	17.20 (8.72)
Duration		
Shortest conversation length	0:00:08	-
Longest conversation length	0:17:00	-
Average conversation length	-	00:03:02 (00:03:40)
Median conversation length	0:05:49	-
Initiation		
AI Coach initiated conversation (rounds)	205	6.83 (3.72)
Participant initiated conversation (rounds)	311	10.42 (7.13)

SD: Standard Deviation

A total of 3,081 messages were exchanged between the participants and the AI Coach, including the AI Coach's default daily message, resulting in 516 rounds of conversation. A conversation was only counted as a round if the participant responded within 30 minutes of receiving the AI Coach's message. Of the total conversations, the majority were initiated by the participants (60.27%) compared to those initiated by the AI Coach. The average number of conversation rounds per participant was 17.20 (SD = 8.72). The longest conversation lasted 17 minutes, while the shortest was 8 seconds. On average, each conversation lasted 3 minutes and 2 seconds (SD = 3 minutes and 40 seconds), with the median conversation duration being approximately 6 minutes (00:05:49).

At the end of week 4, the participants' average SUS score was 85.75 (SD = 3.99) (Table 4), which is considered to reflect excellent usability [47].

Table 4. Scores for the System Usability Scale (SUS) following the intervention

Questions*	M
I think I would like to use the app frequently	3
I found the system to be unnecessaril	1

y complex	
I thought the system was easy to use	4
I think that I would need the support of a technical pers	1

on to be able to use the system	
I found the various functions in the system were well-integrated	3
I thou	2

ght there was too much inconsistency in the system I would imagine that most people would learn to us	4

e the system very quickly	
I found the system very cumbersome to use	1
I felt very confident using the	4

system I needed to learn a lot of things before I could get going with the system	1
System Usability Scal	8

e to tal sc or e (o ut of 10 0)

Based on the average score, participants generally agreed or strongly agreed that the system was easy to use, that they felt confident using it, and that most would learn to use it quickly. Additionally, the average scores indicated that participants generally disagreed or strongly disagreed with the statements suggesting they needed the support of a technical person to use the system and that they needed to learn extensively before being able to use it.

In terms of the user experience questionnaire, 93.33% of participants reported feeling that the AI Coach made them more active, 86.67% indicated that they would recommend the AI Coach to others, and 90.00% expressed a willingness to continue using the AI Coach.

Efficacy

As shown in Table 5 and Figures 3 and Figure 4, from baseline to week 8, the IG showed a significant increase in PA. Specifically, the average daily step count increased by 557 steps and the duration of weekly MVPA increased by 8.73 minutes, both of which reached statistical significance.

Table 5. Average Daily Steps and Weekly MVPA of Groups

Behavioral Performance Indicators	Groups	Pre-intervention (Mean ± SD)	Post-intervention (Mean ± SD)
Steps (steps/day)	CG	7202 ± 808	7094 ± 556
	IG	7427 ± 440	7984 ± 812
MVPA (min/week)	CG	88.23 ± 12.15	89.17 ± 7.64
	IG	92.20 ± 6.08	100.93 ± 9.57

Figure 3. Changes in average daily step count from pre- to post-intervention in the CG and IG

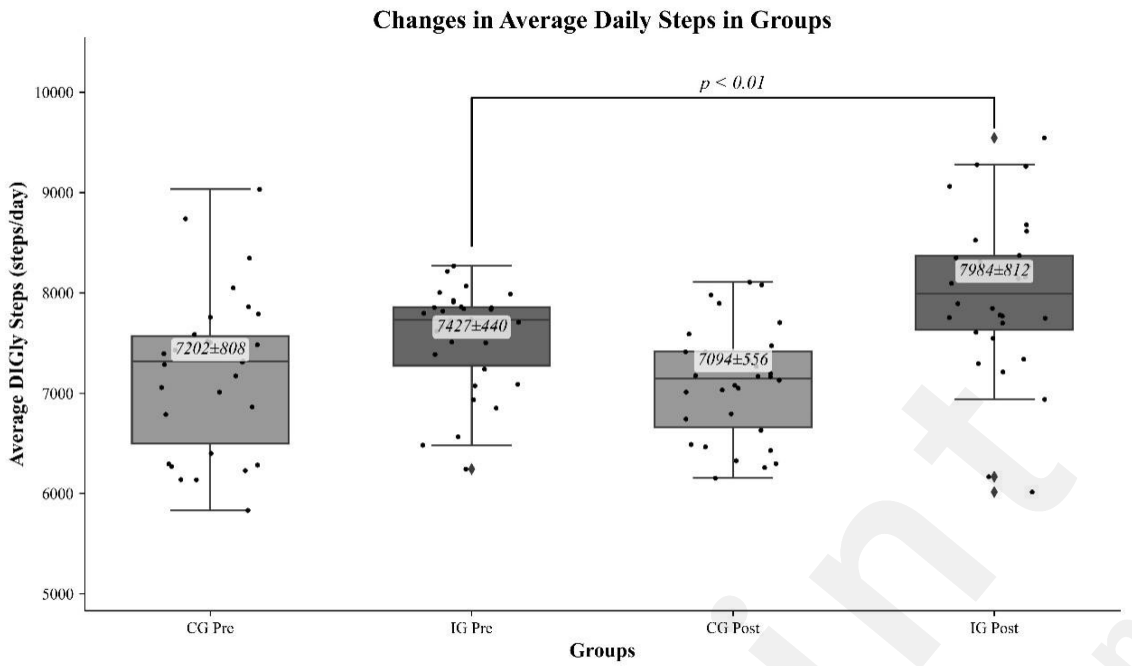
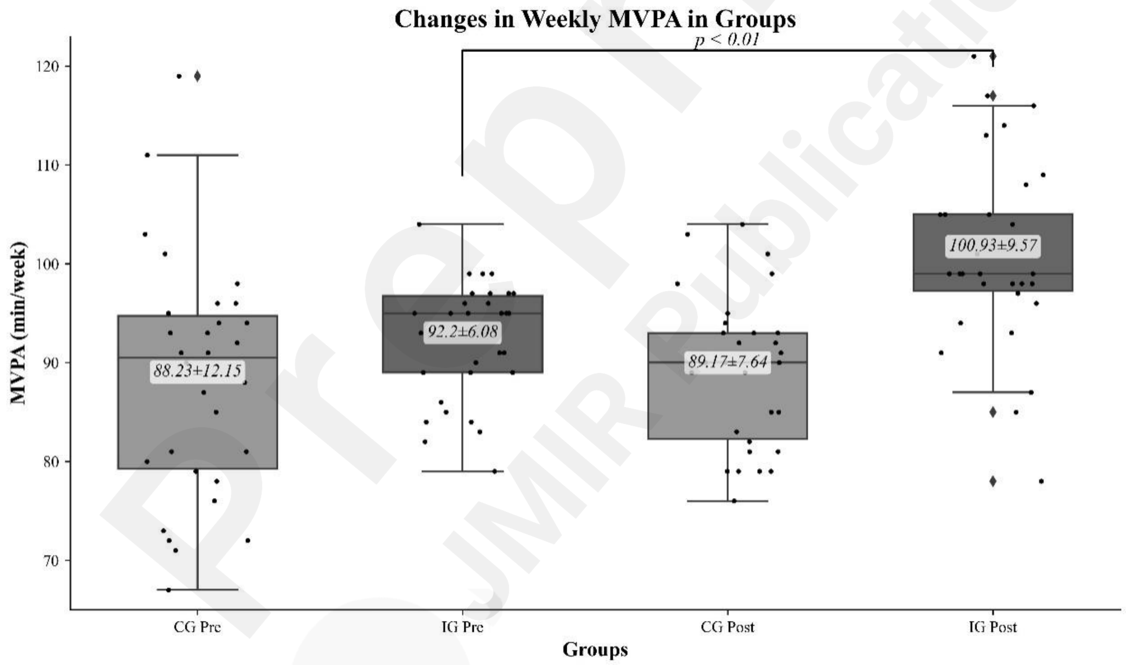


Figure 4. Changes in weekly MVPA from pre- to post-intervention in the CG and IG



Before the intervention, there were no statistically significant differences between the IG and the CG in terms of step count (p = 0.062 or MVPA (p = 0.117) (see Table 6). Following the intervention, the IG exhibited a significant increase in both step count (p < 0.001) and MVPA (p < 0.001), indicating the efficacy of the AI coaching intervention in promoting PA (see Table 6 and Figure. 3 – 4).

Table 6. Comparison of Steps count and MVPA Between Groups

Behavioral Performance Indicators	Comparison	Test Method	p	Levene_p
Steps	CG Pre vs Post	Paired t-test	0.345	
	IG Pre vs Post	Wilcoxon signed-rank test	0.004**	

	CG vs IG (Pre)	Mann–Whitney U test	0.062	0.0545
	CG vs IG (Post)	Unpaired t-test	0.000**	0.1295
MVPA	CG Pre vs Post	Paired t-test	0.638	
	IG Pre vs Post	Paired t-test	0.000**	
	CG vs IG (Pre)	Welch's ANOVA	0.117	0.005*
	CG vs IG (Post)	Welch's ANOVA	0.000**	0.55

**p < 0.01

In contrast, no significant changes were observed in the CG between the baseline and post-intervention assessments for either step count (p = 0.345) or MVPA (p = 0.638), suggesting a lack of an effect in the absence of intervention.

Before the intervention, there were no significant differences between the groups in step count (p = 0.060) or MVPA (p = 0.117). However, post-intervention comparisons revealed that the IG significantly outperformed the CG in both step count and MVPA, with both outcomes reaching statistical significance (p < 0.001 for both). These findings further underscore the positive impact of the AI coaching intervention on PA levels.

Within-group comparisons indicated that increases in step count and MVPA in the IG were statistically significant, as assessed using the Wilcoxon signed-rank test (p < 0.001 for steps count; p < 0.001 for MVPA). In contrast, CG did not exhibit any significant within-group changes. Independent-sample t-tests conducted on post-intervention data showed that the IG had significantly higher step counts and MVPA than the CG (p < 0.001 for both), highlighting a clear intervention effect.

Within-group comparisons indicated that increases in step count and MVPA in the IG were statistically significant, as assessed using the Wilcoxon signed-rank test (p < 0.001 for steps count; p < 0.001 for MVPA). In contrast, CG did not exhibit any significant within-group changes. Independent-sample t-tests conducted on post-intervention data showed that the IG had significantly higher step counts and MVPA than the CG (p < 0.001 for both), highlighting a clear intervention effect.

Adverse Events

No adverse reactions were reported during the intervention.

Discussion

Engagement, Usability and Efficacy

The AI coach has demonstrated the potential to enhance PA levels among university students, proving to be a safe, usable, efficacious, and practical intervention.

The average number of conversation turns with the AI coach was 17.20 ± 8.72, exceeding the 12 ± 8.84 turns reported for Tess during its longer 12-week intervention period [30]. Additionally, unlike Tess [30], where only 26.39% of conversations were initiated by participants, in our study, a significant majority (60.27%) of conversations were initiated by the participants themselves. This suggests that participants were more actively engaged in their interactions with the AI coach, possibly because the AI provided personalized encouragement and practical plans that helped improve their PA levels, better meeting their needs and motivating them to engage in conversations.

Conversation durations ranged from 8 seconds to 17 minutes, which may have depended on the complexity of the questions asked by participants or the number of topics discussed. Longer conversation durations might indicate greater participant interest in certain topics or the need for more detailed information and advice, while shorter durations may suggest simpler inquiries or that participants were satisfied with the responses provided.

The average System Usability Scale (SUS) score was 85.75, which indicates excellent usability, surpassing the scores of previous studies (84 and 75) [48,49]. This could be attributed to the fact that participants were already familiar with using WeChat, and the interface and interaction style of the AI coach resembled that of WeChat, eliminating the need for additional downloads and registration. This reduced barriers to use and increased user acceptance and convenience. Participants rated the usability aspects of the system highly, particularly for ease of use (SUS items 2, 3, 4, 7, 8, 10). However, there were higher expectations regarding the integration of system functions (SUS item 5) and consistency (SUS item 6). This suggests that future improvements should focus on optimizing the integration and consistency of the system’s functionalities to further enhance the user experience.

Before the intervention, no statistically significant differences were found between the IG and CG groups in terms of age, gender, height, weight, or BMI (see Table 2). This baseline equivalence suggests that the two groups were comparable prior to the intervention, providing a reliable foundation for interpreting the differences observed after the intervention.

After 8 weeks of intervention, participants showed a significant increase in their average daily step count (an increase of 393 steps) and weekly MVPA duration (an increase of 4.18 minutes). These results are consistent with those of a previous study, which showed that chatbots combined with wearable devices were more efficacious in enhancing PA levels than using wearable devices alone [40].

Strengths and Limitations

This study is the first to investigate the use of Doubao-based interventions for improving PA levels. Since participants in this study were already using wearable devices, we were able to assess the impact of the chatbot independently of the wearable device, providing a more accurate measure of its effects on PA promotion. This design offers deeper insight into the potential role of chatbots in health behavior change, particularly in their ability to function both in conjunction with existing health management tools and as standalone interventions. Furthermore, the comprehensive data on engagement, usability, and intervention efficacy reported in this study provide a valuable reference for future research and comparisons. The study introduces novel perspectives and methodologies, contributing to the advancement of related research.

However, there are several limitations. First, the relatively small sample size may limit the representativeness and reliability of the findings, which could affect their generalizability. Second, participants' experiences were assessed exclusively through a questionnaire, which may restrict the depth of the investigation and potentially fail to fully capture their actual user experience. Third, while the validity of step count measurements from wearable devices is well established, the estimation of MVPA by consumer-grade devices, although improved in recent years, still exhibits systematic bias and may not precisely reflect true MVPA levels. Therefore, outcomes involving MVPA duration should be interpreted with caution in light of potential measurement limitations.

Future Research and Applications

Future improvements and explorations should address both technological advancements and research methodologies.

From a technological perspective, real-time data acquisition could be automated (as this study manually collected data the following day) to enhance the timeliness and accuracy of data collection. Beyond steps and MVPA, future studies could incorporate additional data types, such as sleep, sedentary behavior, blood pressure, blood glucose, and heart rate, as well as real-time local weather, environmental conditions, and other contextual factors. Integrating voice, images, and video technologies could simulate a more realistic coaching experience, offering participants a richer and more personalized interaction. Furthermore, the system could be individualized based on personal data, exercise goals, and other health objectives, allowing users to customize the AI Coach’s settings to their preferences (e.g., gender, age, communication style), thus enhancing the overall user experience.

In terms of research methodology, increasing the sample size would improve the representativeness and reliability of the findings. Qualitative research should also be incorporated to gain deeper insights into participants' experiences and to provide more detailed and nuanced data to support future intervention improvements. Additionally, expanding the study population beyond university students to include diverse demographics would allow for a more comprehensive evaluation of the AI Coach’s efficacy across different populations, thereby assessing its broader applicability.

Conclusions

The Doubao-based AI Coach has demonstrated significant potential in enhancing PA levels among university students, representing a safe, efficacious, and practical intervention. Future studies could integrate more personalized data, incorporate additional environmental factors, and refine the interaction format to enhance the overall user experience. Moreover, further in-depth research, based on larger sample sizes, is essential to assess the intervention’s efficacy, explore participants’ experiences in greater detail, and validate its applicability across diverse populations, thereby ensuring the generalizability of the results.

References

1. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, Carty C, Chaput J-P, Chastin S, Chou R, Dempsey PC, DiPietro L, Ekelund U, Firth J, Friedenreich CM, Garcia L, Gichu M, Jago R, Katzmarzyk PT, Lambert E, Leitzmann M, Milton K, Ortega FB, Ranasinghe C, Stamatakis E, Tiedemann A, Troiano RP, van der Ploeg HP, Wari V, Willumsen JF. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec;54(24):1451–1462. PMID:33239350

2. Kraus WE, Janz KF, Powell KE, Campbell WW, Jakicic JM, Troiano RP, Sprow K, Torres A, Piercy KL. Daily Step Counts for Measuring Physical Activity Exposure and Its Relation to Health. Med Sci Sports Exerc 2019 Jun;51(6):1206–1212. PMID:31095077

3. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, Galuska DA,

Jefferis BJ, Kraus WE, Lee I-M, Matthews CE, Omura JD, Patel AV, Pieper CF, Rees-Punia E, Dallmeier D, Klenk J, Whincup PH, Dooley EE, Pettee Gabriel K, Palta P, Pompeii LA, Chernofsky A, Larson MG, Vasan RS, Spartano N, Ballin M, Nordström P, Nordström A, Anderssen SA, Hansen BH, Cochrane JA, Dwyer T, Wang J, Ferrucci L, Liu F, Schrack J, Urbanek J, Saint-Maurice PF, Yamamoto N, Yoshitake Y, Newton RL, Yang S, Shiroma EJ, Fulton JE. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022 Mar;7(3):e219–e228. doi: 10.1016/S2468-2667(21)00302-9

4. Aune D, Norat T, Leitzmann M, Tonstad S, Vatten LJ. Physical activity and the risk of type 2 diabetes: a systematic review and dose–response meta-analysis. Eur J Epidemiol 2015 Jul 1;30(7):529–542. doi: 10.1007/s10654-015-0056-z

5. McMahon EM, Corcoran P, O'Regan G, Keeley H, Cannon M, Carli V, Wasserman C, Hadlaczky G, Sarchiapone M, Apter A, Balazs J, Balint M, Bobes J, Brunner R, Cozman D, Haring C, Iosue M, Kaess M, Kahn J-P, Nemes B, Podlogar T, Poštuvan V, Sáiz P, Sisask M, Tubiana A, Värnik P, Hoven CW, Wasserman D. Physical activity in European adolescents and associations with anxiety, depression and well-being. Eur Child Adolesc Psychiatry 2017 Jan 1;26(1):111–122. doi: 10.1007/s00787-016-0875-9

6. Moore SC, Lee I-M, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, Keadle SK, Arem H, Berrington de Gonzalez A, Hartge P, Adami H-O, Blair CK, Borch KB, Boyd E, Check DP, Fournier A, Freedman ND, Gunter M, Johannson M, Khaw K-T, Linet MS, Orsini N, Park Y, Riboli E, Robien K, Schairer C, Sesso H, Spriggs M, Van Dusen R, Wolk A, Matthews CE, Patel AV. Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. JAMA Intern Med 2016 Jun 1;176(6):816–825. doi: 10.1001/jamainternmed.2016.1548

7. Wanner M, Richard A, Martin B, Faeh D, Rohrmann S. Associations between self-reported and objectively measured physical activity, sedentary behavior and overweight/obesity in NHANES 2003–2006. Int J Obes Nature Publishing Group; 2017 Jan;41(1):186–193. doi: 10.1038/ijo.2016.168

8. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health Elsevier; 2018 Oct 1;6(10):e1077–e1086. PMID:30193830

9. Irwin JD. The prevalence of physical activity maintenance in a sample of university students: a longitudinal study. J Am Coll Health J ACH 2007;56(1):37–41. PMID:17711824

10. Mohammed G, Md Said S, Azuhairi, Ariffin A, Kamaruzaman J. Physical Inactivity and Its Associated Factors among University Students. IOSR J Dent Med Sci 2014;13(10):119–130. doi: 10.9790/0853-13101119130

11. Yıldız ME, Günel İ, Dalbudak İ. The relationship between physical activity and mindful awareness of university students. Phys Educ Stud 2024 Aug 30;28(4):234–241. doi: 10.15561/20755279.2024.0406

12. Chen B, Zhou Y, Huang Z. Associations of physical activity, sleep quality with anxiety and depressive symptoms among college students. Chin J Sch Health 2024;45(5):684–688. doi: 10.16835/j.cnki.1000-9817.2024139

13. Manuela P, Gheorghe PA. REPORT ON THE LEVEL OF PHYSICAL ACTIVITY AMONG STUDENTS. 2020;(2).

14. Sánchez-Herrera S, Cubero J, Feu S, Durán-Vinagre MÁ. Motivation Regarding Physical Exercise among Health Science University Students. Int J Environ Res Public Health Multidisciplinary Digital Publishing Institute; 2022 Jan;19(11):6524. doi: 10.3390/ijerph19116524

15. Dietz P, Watson ED, van Poppel M, Tibubos AN. Editorial: Public health promotion in university students. Front Public Health Frontiers; 2022 Aug 22;10. doi: 10.3389/fpubh.2022.993982

16. Chao D-P. Health-promoting lifestyle and its predictors among health-related and non-health-related university students in Taiwan: a cross-sectional quantitative study. BMC Public Health 2023 May 5;23(1):827. doi: 10.1186/s12889-023-15760-2

17. Zhao Y, Yin Y, Li J. Research Progress on Physical Activity Among College Students Based on Socio-Ecological Models. Contemp Sports Technol 2024;14(30):153–157. doi: 10.16655/j.cnki.2095-2813.2024.30.039

18. Plotnikoff RC, Costigan SA, Williams RL, Hutchesson MJ, Kennedy SG, Robards SL, Allen J, Collins CE, Callister R, Germov J. Effectiveness of interventions targeting physical activity, nutrition and healthy weight for university and college students: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2015 Apr 1;12(1):45. doi: 10.1186/s12966-015-0203-7

19. Grassini S, Aasen ML, Møgelvang A. Understanding University Students' Acceptance of ChatGPT: Insights from the UTAUT2 Model. Appl Artif Intell Taylor & Francis; 2024 Dec 31; Available from: https://www.tandfonline.com/doi/abs/10.1080/08839514.2024.2371168 [accessed Jul 26, 2025]

20. Wang Y, Zhang X, Wang L. Assessing the Intention to Use Sports Bracelets Among Chinese University Students: An Extension of Technology Acceptance Model With Sports Motivation. Front Psychol Frontiers; 2022 Mar 23;13. doi: 10.3389/fpsyg.2022.846594

21. Longhini J, Marzaro C, Bargeri S, Palese A, Dell'Isola A, Turolla A, Pillastrini P, Battista S, Castellini G, Cook C, Gianola S, Rossettini G. Wearable Devices to Improve Physical Activity and Reduce Sedentary Behaviour: An Umbrella Review. SPORTS Med-OPEN New York: Springer; 2024 Jan 14;10(1):9. doi: 10.1186/s40798-024-00678-9

22. 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 Toronto: Jmir Publications, Inc; 2023 Feb 24;25:e40789. doi: 10.2196/40789

23. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, Surian D, Gallego B, Magrabi F, Lau AY. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc Oxford University Press; 2018;25(9):1248–1258.

24. Maher CA, Davis CR, Curtis RG, Short CE, Murphy KJ. A Physical Activity and Diet Program Delivered by Artificially Intelligent Virtual Health Coach: Proof-of-Concept Study. JMIR MHealth UHealth 2020 Jul 10;8(7):e17558. doi: 10.2196/17558

25. Oh YJ, Zhang J, Fang M-L, Fukuoka Y. 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(1):160. doi: 10.1186/s12966-021-01224-6

26. Kocielnik R, Hsieh G. Facilitating Self-learning in Behavior Change Through Long-term Intelligent Conversational Assistance. 23rd Int Conf Intell User Interfaces 2018. p. 683–684.

27. Kramer J-N, Künzler F, Mishra V, Presset B, Kotz D, Smith S, Scholz U, Kowatsch T. Investigating Intervention Components and Exploring States of Receptivity for a Smartphone App to Promote Physical Activity: Protocol of a Microrandomized Trial. JMIR Res Protoc 2019 Jan 31;8(1):e11540. PMID:30702430

28. Künzler F, Mishra V, Kramer J-N, Kotz D, Fleisch E, Kowatsch T. Exploring the State-of-Receptivity for mHealth Interventions. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019;3(4):140:1-140:27. doi: 10.1145/3369805

29. Piao M, Kim J, Ryu H, Lee H. Development and Usability Evaluation of a Healthy Lifestyle Coaching Chatbot Using a Habit Formation Model. Healthc Inform Res 2020 Oct 1;26(4):255–264. PMID:33190459

30. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440–447. doi: 10.1093/tbm/ibz043

31. Singh B, Olds T, Brinsley J, Dumuid D, Virgara R, Matricciani L, Watson A, Szeto K, Eglitis E, Miatke A, Simpson CEM, Vandelanotte C, Maher C. Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours. Npj Digit Med Nature Publishing Group; 2023 Jun 23;6(1):1–10. doi: 10.1038/s41746-023-00856-1

32. Abdullah M, Madain A, Jararweh Y. ChatGPT: Fundamentals, Applications and Social Impacts. 2022 Ninth Int Conf Soc Netw Anal Manag Secur SNAMS Milan, Italy: IEEE; 2022. p. 1–8. doi: 10.1109/SNAMS58071.2022.10062688

33. Gole M, Nwadiugwu W-P, Miranskyy A. On Sarcasm Detection with OpenAI GPT-Based Models. 2024 34th Int Conf Collab Adv Softw Comput CASCON Toronto, ON, Canada: IEEE; 2024. p. 1–6. doi: 10.1109/CASCON62161.2024.10837875

34. OpenAI. ChatGPT. ChatGPT. 2024. Available from: https://openai.com/chatgpt/ [accessed Aug 29, 2024]

35. Zhang B, Ding D, Jing L, Dai G, Yin N. How would Stance Detection Techniques Evolve after the Launch of ChatGPT? arXiv; 2024. doi: 10.48550/arXiv.2212.14548

36. Doubao. Doubao 1.5pro - Doubao Team. 2025. Available from: https://team.doubao.com/en/special/doubao_1_5_pro?view_from=research [accessed Mar 22, 2025]

37. Grok-2 Beta Release | xAI. Available from: https://x.ai/news/grok-2 [accessed Mar 22, 2025]

38. Introducing Gemini: our largest and most capable AI model. Google. 2023. Available from: https://blog.google/technology/ai/google-gemini-ai/ [accessed Mar 22, 2025]

39. Team K, Du A, Gao B, Xing B, Jiang C, Chen C, Li C, Xiao C, Du C, Liao C, Tang C, Wang C, Zhang D, Yuan E, Lu E, Tang F, Sung F, Wei G, Lai G, Guo H, Zhu H, Ding H, Hu H, Yang H, Zhang H, Yao H, Zhao H, Lu H, Li H, Yu H, Gao H, Zheng H, Yuan H, Chen J, Guo J, Su J, Wang J, Zhao J, Zhang J, Liu J, Yan J, Wu J, Shi L, Ye L, Yu L, Dong M, Zhang N, Ma N, Pan Q, Gong Q, Liu S, Ma S, Wei S, Cao S, Huang S, Jiang T, Gao W, Xiong W, He W, Huang W, Wu W, He W, Wei X, Jia X, Wu X, Xu X, Zu X, Zhou X, Pan X, Charles Y, Li Y, Hu Y, Liu Y, Chen Y, Wang Y, Liu Y, Qin Y, Liu Y, Yang Y, Bao Y, Du Y, Wu Y, Wang Y, Zhou Z, Wang Z, Li Z, Zhu Z, Zhang Z, Wang Z, Yang Z, Huang Z, Huang Z, Xu Z, Yang Z. Kimi k1.5: Scaling Reinforcement Learning with LLMs. arXiv; 2025. doi: 10.48550/arXiv.2501.12599

40. Hassoon A, Baig Y, Naiman DQ, Celentano DD, Lansey D, Stearns V, Coresh J, Schrack J, Martin SS, Yeh H-C, Zeilberger H, Appel LJ. Randomized trial of two artificial intelligence coaching interventions to increase physical activity in cancer survivors. Npj Digit Med Nature Publishing Group; 2021 Dec 9;4(1):1–6. doi: 10.1038/s41746-021-00539-9

41. News Assistant. The evaluation results of Doubao large model have been disclosed, showing a 19% improvement compared with the previous generation “Alauda.” Synced Mach. 2024. Available from: https://www.jiqizhixin.com/articles/2024-05-24-11 [accessed Apr 18, 2025]

42. Degroote L, De Bourdeaudhuij I, Verloigne M, Poppe L, Crombez G. The Accuracy of Smart Devices for Measuring Physical Activity in Daily Life: Validation Study. JMIR MHEALTH UHEALTH Toronto: Jmir Publications, Inc; 2018 Dec;6(12):e10972. doi: 10.2196/10972

43. Degroote L, Hamerlinck G, Poels K, Maher C, Crombez G, De Bourdeaudhuij I, Vandendriessche A, Curtis RG, DeSmet A. Low-Cost Consumer-Based Trackers to Measure Physical Activity and Sleep Duration Among Adults in Free-Living Conditions: Validation Study. JMIR MHealth UHealth 2020 May 19;8(5):e16674. PMID:32282332

44. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, Buote R, Heerden DV, Luan H, Cullen K, Slade L, Taylor NGA. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. JMIR MHealth UHealth 2020 Sep 8;8(9):e18694. doi: 10.2196/18694

45. Michie S, Atkins L, West R. The behaviour change wheel: a guide to designing interventions. Sutton: Silverback Publishing; 2014. ISBN:978-1-912141-08-1

46. Gooden C. (8) TRACE: A Framework for Prompt Engineering for Generative AI in Education | LinkedIn. TRACE Framew Prompt Eng Gener AI Educ. 2023. Available from: https://www.linkedin.com/pulse/trace-framework-prompt-engineering-generative-ai-education-gooden/ [accessed Jul 28, 2024]

47. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Intl J Human–Computer Interact Taylor & Francis Group; 2008 Jul 29; doi: 10.1080/10447310802205776

48. Daly RM, Gianoudis J, Hall T, Mundell NL, Maddison R. Feasibility, Usability, and Enjoyment of a Home-Based Exercise Program Delivered via an Exercise App for Musculoskeletal Health in Community-Dwelling Older Adults: Short-term Prospective Pilot Study. JMIR MHealth UHealth 2021 Jan 13;9(1):e21094. doi: 10.2196/21094

49. Jansons P, Dalla Via J, Daly RM, Fyfe JJ, Gvozdenko E, Scott D. Delivery of Home-Based

Exercise Interventions in Older Adults Facilitated by Amazon Alexa: A 12-week Feasibility Trial. J Nutr Health Aging 2022 Jan 1;26(1):96–102. doi: 10.1007/s12603-021-1717-0