GPT-4 as a Virtual Fitness Coach: An Evaluation of Its Effectiveness in Providing Weight Loss and Fitness Guidance

Guodong Zhang, Guanchong Li, Hansen Li, Yuqin Su, Yun Li

Submitted to: Journal of Medical Internet Research
on: August 16, 2024

Table of Contents

GPT-4 as a Virtual Fitness Coach: An Evaluation of Its Effectiveness in Providing Weight Loss and Fitness Guidance

Guodong Zhang[1] A; Guanchong Li[2] A; Hansen Li[2] A2; Yuqin Su[2] AA; Yun Li[1] AAA

[1]College of Physical Education, Southwest University Chongqing CN
[2]Institute of Sport Science, College of Physical Education, Southwest University Chongqing CN

Corresponding Author:
Guodong Zhang A
College of Physical Education, Southwest University
College of Physical Education, Southwest University, Beibei District, Chongqing City, China
Chongqing
CN

Abstract

Background: With obesity posing an increasing public health challenge, the demand for personalized weight loss and fitness solutions has intensified. Given advancements in artificial intelligence, like GPT-4, this study evaluates its effectiveness as a virtual fitness coach for creating personalized plans.

Method: We selected a 24-year-old female from Chongqing, China, providing detailed personal information. Using this, both GPT-4 and three local professional coaches formulated 16-week fitness plans. Senior health science experts evaluated these plans across four dimensions: personalization, effectiveness, comprehensiveness, and safety. Descriptive and inferential statistics were employed for analysis.

Results: GPT-4 excelled in personalization (M=12.80, SD=0.84) compared to coaches (M=11.53, SD=0.46). However, coaches slightly outperformed in effectiveness (GPT-4: M=12.60; Coaches: M=12.80), safety (GPT-4: M=12.20; Coaches: M=12.33), and comprehensiveness (GPT-4: M=12.00; Coaches: M=12.13), with no significant differences (P > .05). These findings highlight comparable performance but suggest potential for GPT-4 in personalized exercise prescriptions.

Conclusions: GPT-4 shows promise as a virtual fitness coach for personalized weight loss and fitness guidance. Yet, due to technological constraints, it cannot fully replace human coaches. Future research should explore enhancing AI models' applicability in sports and their collaboration with coaches for optimal personalized fitness solutions.

(JMIR Preprints 16/08/2024:65470)
DOI: https://doi.org/10.2196/preprints.65470

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

Original Manuscript

GPT-4 as a Virtual Fitness Coach: An Evaluation of Its Effectiveness in Providing Weight Loss and Fitness Guidance

ABSTRACT

Background: With obesity posing an increasing public health challenge, the demand for personalized weight loss and fitness solutions has intensified. Given advancements in artificial intelligence, like GPT-4, this study evaluates its effectiveness as a virtual fitness coach for creating personalized plans.

Method: We selected a 24-year-old female from Chongqing, China, providing detailed personal information. Using this, both GPT-4 and three local professional coaches formulated 16-week fitness plans. Senior health science experts evaluated these plans across four dimensions: personalization, effectiveness, comprehensiveness, and safety. Descriptive and inferential statistics were employed for analysis.

Results: GPT-4 excelled in personalization (M=12.80, SD=0.84) compared to coaches (M=11.53, SD=0.46). However, coaches slightly outperformed in effectiveness (GPT-4: M=12.60; Coaches: M=12.80), safety (GPT-4: M=12.20; Coaches: M=12.33), and comprehensiveness (GPT-4: M=12.00; Coaches: M=12.13), with no significant differences ($P > .05$). These findings highlight comparable performance but suggest potential for GPT-4 in personalized exercise prescriptions.

Conclusions: GPT-4 shows promise as a virtual fitness coach for personalized weight loss and fitness guidance. Yet, due to technological constraints, it cannot fully replace human coaches. Future research should explore enhancing AI models' applicability in sports and their collaboration with coaches for optimal personalized fitness solutions.

Keywords: GPT-4, virtual Fitness coach, personalized fitness solutions, technological limitations, health sciences.

1 Introduction

Obesity has become a significant public health challenge in contemporary society (Sabin, Marini, & Nosek, 2012; Schwartz, Chambliss, Brownell, Blair, & Billington, 2003). Research predicts that by 2035, approximately 1.9 billion people worldwide will be facing obesity issues, potentially making it one of the most significant health threats of the 21st century (James, 2018; Štempeľová, Takáč, Hudáková, & Researches, 2023). On the other hand, obesity has been proven to be detrimental to both physical and psychological health. The World Health Organization warns that the rising rates of obesity will increase the risks of various diseases, including heart disease, diabetes, and cancer (Almeida, Dera, Murphy, & Santosa, 2024; Bouchard, Depres, & Tremblay, 1993; Iqbal, Masood, & Ikram, 2019).

Currently, there are a variety of treatment options for obesity including pharmacological interventions (Blüher, 2008) and surgical procedures (Lim, 2024; Slomski, 2022). However, these methods are considered to carry associated health risks and may not provide long-term control of obesity (Derickson et al., 2018; Shi et al., 2024) In this context, behavioral therapy is the most advocated approach, primarily involving dietary adjustments and increased physical activity (Villareal et al., 2011; Zhi, 2022). To date, countless studies have demonstrated that physical exercise plays a crucial role in controlling weight and improving health (McInnis, 2000; Petridou, Siopi, & Mougios, 2019). However, in a wide range of physical activities and exercises, we are more concerned with maximizing the benefits of exercise while minimizing the risks. Theoretically, inappropriate training methods could lead to reduced benefits or increased risks of exercise, issues that have already garnered attention in sports training. Considering the differences among individuals, personalized scientific planning is crucial. Buford (2013) posits that personalized exercise programs are key in addressing obesity and maintaining a healthy lifestyle. This customized approach takes into account individual characteristics and needs, thereby avoiding the negative impacts that may arise from non-customized exercise plans (Buford, Roberts, & Church, 2013).

For the average individual, there is a lack of knowledge in training science, making it challenging to design appropriate plans for themselves. Although fitness coaches can offer professional advice, their high costs and the inconsistency in professional quality make widespread accessibility difficult to achieve. As mentioned in Melmer's (2018) study, patients with heart disease require personalized exercise guidance. In such cases, the involvement of professional coaches is critically important. However, due to cost and accessibility issues, such services are not widely available (Galiuto, Fedele, Vitale, & Lucini, 2019). Especially in developing countries like China, the demand brought by the massive population does not match the relatively underdeveloped community sports service resources, which further reduces the opportunities for the general public to obtain high-quality fitness guidance (Ren, Liu, & Health, 2021).

Against this backdrop, as an emerging natural language processing tool, GPT (especially the latest version, GPT-4) demonstrates substantial potential (Arslan, 2023). It shows promise in providing cost-effective, user-friendly personalized health and fitness advice. Research by Hirak Mazumdar, K. Bhattarai, and others suggests that personalized health services are becoming a reality (Bhattarai et al., 2023; Mazumdar, Chakraborty, Sathvik, Panigrahi, & Informatics, 2023). This could potentially transform traditional fitness models and drive innovation in the fitness industry. Current research underscores the potential application of GPT in the field of exercise science. For instance, Wang M et al. (2023) summarized the performance of Chat GPT across various domains in their research, stating that its performance in the field of exercise science (Szabo, 2023) leaves much to be desired. However, there is a notable lack of in-depth research into the development and evaluation of personalized training programs. Vogel's study (2009) explored the ability of GPT to identify individual preferences for exercise plans, but did not thoroughly evaluate the actual effectiveness of these plans (Vogel et al., 2009). Although the research by Washif J et al. (2023) confirmed that GPT-4 has improved in designing personalized resistance training programs compared to its predecessors, this improvement was only observed in comparisons between artificial intelligence (Washif, Pagaduan, James, Dergaa, & Beaven, 2024). Our research aims to fill this gap by evaluating whether GPT-4 can replace or assist traditional fitness coaches, and by providing safe and effective exercise guidance for a wide range of users.

2 Methods

2.1Test Samples

The rising incidence of obesity among young populations has attracted widespread attention (M. Ng et al., 2014). Considering the research resources, we selected a female volunteer with a willingness to lose weight from a weight loss group at a university. Her physical activity was insufficient due to academic pressure, and her physical indicators were close to the average level of women in the weight loss group at the same university, making her the focus of our study. Based on preliminary surveys, her willingness to lose weight is similar to that of most women, enabling her to represent, to some extent, the young female population facing similar health challenges.

Table 1. Basic Information of Test Sample

Parameter	Test Value	Parameter	Test Value
Height	159 cm	BMI	25.1
Weight	63.5 kg		
Standard Weight (Reference)	49KG	Body Fat Percentage	30.24%

Waist Circumference	86CM	Waist-to-Hip Ratio	0.86
Hip Circumference	100CM		
Degree of Obesity (%)	13.5%	Basal Metabolic Rate	1701kcal
Resting Heart Rate	74bpm	Obesity Type	Overweight

Based on the screening results from the Physical Activity Readiness Questionnaire (PAR-Q) (Crookham, 2013), an exercise prescription was subsequently formulated.

2.2 Study Design

We used expert evaluations as the benchmark to compare the differences between GPT and fitness coaches in generating fitness plans, in order to determine whether GPT can be used as a substitute for human fitness coaches.

2.2.1 Selection of Experts

The selection of experts was based on their extensive experience and professional knowledge in the fields of sports, fitness coaching, and related medical disciplines. Through a collaborative approach, we integrated the expertise of physicians, exercise scientists, and fitness coaches to jointly evaluate the exercise prescriptions (Zhu et al., 2024). We invited senior experts XXH and ZJ from the China Fitness and Bodybuilding Association, trainers PGF and SXB from the School of Physical Education at Chongqing University of Posts and Telecommunications, and ZZW from the Health Technology Department of Huawei to participate in the evaluation. These experts not only hold professional fitness certification but also possess years of experience in fitness and weight loss coaching, surpassing the level of service typically available to the general public.

Table 2. Basic Information of Experts

Name (Abbreviation)	Affiliation	Field of Work
XXH	China Fitness and Bodybuilding Association	Sports Medicine
ZJ	China Fitness and Bodybuilding Association	Sports Medicine
SXB	Chongqing University of Posts and Telecommunications	Fitness and Bodybuilding
PGF	Chongqing University of Posts and Telecommunications	Equipment-based Fitness
ZZW	Huawei Health Technology Department	Health Science

2.2.2 Application of GPT

In this study, we utilized the GPT-4 model. Although it is not the latest model from OpenAI, it has been extensively validated and tested, demonstrating exceptional performance across various fields. How about to better verify the specific performance of GDP 4 during weight loss? We have adopted a comprehensive strategy. Firstly, based on detailed body data of a 24-year-old female, including age, height, weight, BMI, and body fat percentage, we utilized the GPT-4 model to generate a 16-week training program. This duration is considered ideal as it ensures effective training outcomes while effectively preventing overtraining and the accumulation of fatigue (Bompa & Buzzichelli, 2019; Issurin, 2010). The program is designed to meet her needs for muscle gain and fat loss, while also improving cardiovascular function and reducing heart rate, with the ultimate goal of enhancing overall physical fitness. Subsequently, we further instructed GPT-4 to enhance the original program by adjusting parameters such as training intensity, rest intervals, and exercise cadence. This was done to ensure the specificity and scientific validity of the program (Geerling, Mateer, Wooten, & Damodaran, 2023; Washif et al., 2024).

Given that GPT-4 tends to stabilize after generating 15 consecutive question-and-answer pairs (Hadi et al., 2023). We will focus on analyzing three independent training plans derived from these 15 interactions. This analysis aims to evaluate their practicality and accuracy in fitness guidance. To ensure a fair and scientific assessment, each professional coach will also provide a corresponding exercise plan for comparison with the GPT-4 generated plans.

Here are the specific prompts we provided to GPT:

Table 3: Prompt Content

Sequence	Content
PROMPT 1	You are a seasoned fitness expert and an ACE-certified personal trainer. Your task is to generate a 16-week weight loss training plan based on the personal physical metrics I provide, aiming to enhance physical fitness, reduce excess body fat, improve cardiovascular function, and lower heart rate. Here are the specific personal details: Female; 24 years old; Height: 159 cm; BMI: 25.1; Weight: 63.5 kg; Body fat percentage: 30.24%; Reference standard weight: 49 kg; Waist circumference: 86 cm; Waist-to-hip ratio: 0.86; Hip circumference: 100 cm; Obesity degree: 13.5%; Basal metabolic rate: 1701 kcal; Resting heart rate: 74 bpm; Obesity type: Overweight.
PROMPT 2	Due to high academic pressure, I have little time for exercise and do not restrict my diet or calorie intake. Currently, I can only train about three times a week. Please personalize the plan accordingly. If you need more detailed information, let me know in advance.
PROMPT 3	Please use the NASM OPT model combined with appropriate aerobic exercises and adjust according to the FITT principles to provide a training plan that can improve my exercise performance. Also, specify the precautions during the workout process.
PROMPT 4	Please include the following information in the training plan: load intensity (percentage), rest intervals, and exercise tempo.
PROMPT 5	Briefly explain the rationale for the recommended plan details or variables in bullet points.
PROMPT 6	Detail the workout content for each day in a concise manner and output it in an EXCEL table format.

(Note: The more detailed the questions posed to GPT, the higher the accuracy of the responses received (Hadi et al., 2023). Continuous follow-up questions were made to refine and perfect the training plan. The above only showcases the main prompts.)

2.2.3 Selection and Consultation of Fitness Coaches

Given that the test subject is a female residing in Beibei District, Chongqing, we hired three of the highest-rated fitness coaches from the top three gyms in this area, based on their user reviews and national certification credentials. These ratings were sourced from Chinese online merchant rating platforms. This selection aims to ensure that the study's findings have strong reference and applicability to the general population in similar urban settings. To maintain comparability, the instructions provided to the fitness coaches were kept consistent with those given to GPT, although such information may be somewhat limited for real-life service provision.

2.3 Evaluation Criteria and Procedure

We employed a single-blind evaluation method, where experts were unaware of whether the plans were generated by GPT or real fitness coaches. The evaluation criteria were primarily based on the FITT model (Frequency, Intensity, Time, Type) (Crookham, 2013; Deng, 2015), a well-established framework for assessing exercise programs, and were structured around the following four key dimensions:

1. Personalization: This dimension examines whether the plan adequately considers the participant's age, gender, current fitness level, existing health issues, personal fitness goals, and exercise preferences. The aim is to evaluate whether the plan can effectively adapt and adjust to individual characteristics.

2. Effectiveness: This dimension focuses on the plan's functionality in achieving the predetermined fitness goals, such as improving cardiovascular health, increasing muscle strength, and enhancing flexibility. It assesses whether the frequency, intensity, time, and type of exercises are scientifically designed to meet individual goals and whether the plan includes a progressive strategy to achieve these goals.

3. Safety: The core of this dimension is whether the plan adheres to safety guidelines and best practices. It evaluates whether the plan considers the individual's physical limitations, injury risks, or special medical conditions and whether it provides correct exercise guidance and techniques to prevent injuries during the workout process.

4. Comprehensiveness: This dimension addresses whether the plan holistically meets the individual's primary fitness goals and overall mental and physical health. It examines whether the plan integrates cardiovascular training, strength training, flexibility training, and neuromotor training. Additionally, it considers factors related to a healthy lifestyle, such as stress management, sleep, and nutrition.

Each dimension includes four sub-criteria, scored on a four-point Likert scale (M. Wang et al., 2023) (1 = Unacceptable, 2 = Acceptable, 3 = Satisfactory, 4 = Very Satisfactory). Each sub-criterion has a maximum score of 4 points, resulting in a maximum score of 16 points per dimension. Consequently, the total score for the entire evaluation system is 64 points (see Table 4).

Table 4 Evaluation Criteria Dimension

Dimension	Sub-criteria	Highly Satisfied	Satisfied	Acceptable	Unacceptable
Personalization	(1) Does it match the individual's age, gender, and current fitness level?				
	(2) Does it take into account the individual's medical history, any existing injuries or limitations, and personal fitness goals?				
	(3) Does it exhibit flexibility to adapt to changes in personal circumstances or preferences over time?				
	(4) Does it consider the individual's exercise preferences and obstacles to sustaining exercise?				
Effectiveness	(1) Can it address the specific physiological adaptability required to achieve personal fitness goals (e.g., improving cardiovascular health, muscle strength, and flexibility)?				
	(2) Analyzes the variables of the exercise plan such as frequency, intensity, time, and type in relation to personal fitness goals and current capabilities.				
	(3) Assesses the progressive plan of the exercise regime, ensuring a gradual and appropriate increase in workout intensity, duration, and complexity.				
	(4) Evaluates whether objective assessments (such as fitness tests, body composition analysis) are incorporated to track and monitor individual progress.				
Safety	(1) Does it comply with recognized safety guidelines and recommended best practices for exercise?				
	(2) Does it take into consideration the individual's				

	current physical limitations, injury risks or medical conditions, and whether the exercise selection and modification are appropriate? （3）Includes correct exercise technique guidance and monitoring to prevent injuries. （4）Considers exercise intensity, duration, and recovery periods to ensure the individual's safety and health.	
Comprehensiveness	（1）Does it not only meet the individual's primary fitness goals but also cater to their overall physical and mental health? （2）Incorporates a balanced approach to fitness, including cardiovascular, muscle, flexibility, and neuromotor training components. （3）Includes stress management techniques such as relaxation, mindfulness, or recovery activities. （4）Considers personal lifestyle factors (such as sleep, nutrition) and their integration with these health aspects.	

2.4 Data Analysis

We confirmed the normality of the data using the Shapiro-Wilk test and the Kolmogorov-Smirnov test. The scoring results are presented as means and standard deviations. Furthermore, we conducted inferential statistical analysis using an independent samples t-test to compare the effectiveness of the two types of fitness plans and to explore the potential capabilities of GPT-4 in the role of a fitness coach. All data analyses were performed using the statistical software SPSS (version 26), with the significance level set at 0.05.

3 Results

This study compared the fitness plans generated by GPT-4 with those created by three professional fitness coaches using both descriptive and inferential statistical analyses. The evaluation focused on four key dimensions: personalization, effectiveness, safety, and comprehensiveness.

3.1 Descriptive Statistics

We used descriptive statistical methods to analyze the average scores of the two groups of fitness plans. The results showed that GPT-4 scored higher in personalization (M=12.80, SD=0.84) compared to the fitness coaches (M=11.53, SD=0.46). However, in terms of effectiveness (GPT-4: M=12.60, Coaches: M=12.80), safety (GPT-4: M=12.20, Coaches: M=12.33), and comprehensiveness (GPT-4: M=12.00, Coaches: M=12.13), the fitness coaches held a slight advantage. Nevertheless, these differences did not reach statistical significance, suggesting that the overall performance of the two types of fitness plans is comparable.

3. 2 Inferential Statistics

To further verify whether the differences between the two groups were statistically significant, we conducted independent samples t-tests. The results are shown below:

Table 5. Results of independent samples t-test

Dimension	GPT-4 Mean	Coaches Mean	t-value	df	p-value	95% Confidence Interval
Personalization	12.80	11.53	−2.38	2	.141	[−3.56，1.03]
Effectiveness	12.60	12.80	0.87	2	.478	[−0.79，1.19]
Comprehensiveness	12.20	12.33	0.14	2	.899	[−3.88，4.14]
Safety	12.00	12.13	0.16	2	.885	[−3.36，3.62]

Although GPT-4 scored higher in personalization compared to the fitness coaches, the difference did not reach statistical significance (P > .05). This indicates that, under the sample and evaluation criteria used in this study, the fitness plans generated by GPT-4 and those created by human coaches exhibit similar performance across the evaluated dimensions.

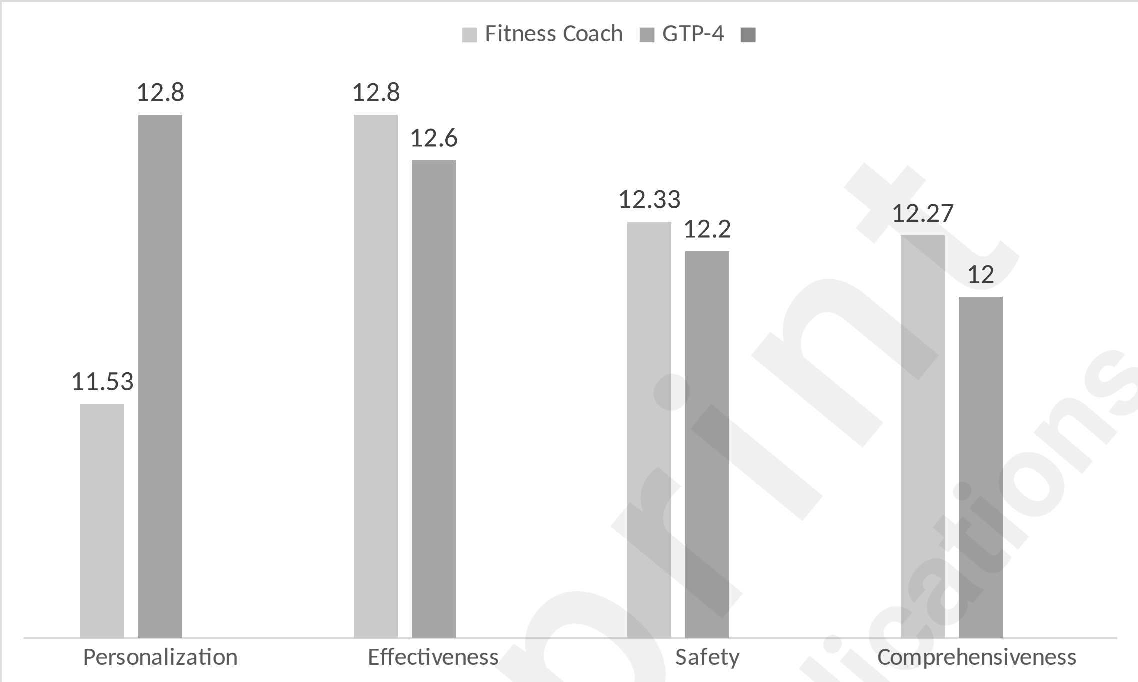

Figure 1. Comparison of Scores Between the Two Groups of Training Plans

These findings suggest that, without considering ongoing feedback and adjustments to the training plans, GPT-4 has potential in creating personalized exercise prescriptions. However, in terms of effectiveness, safety, and comprehensiveness, the plans generated by GPT-4 are comparable to those developed by professional fitness coaches, indicating an overall similar performance between the two.

4 Discussion

This study aimed to examine and compare the ability of the artificial intelligence language model GPT-4 and human fitness coaches in creating personalized fitness plans. The results indicate that while GPT-4 scored slightly higher in the personalization dimension, the difference did not reach statistical significance. In other dimensions, the scores of the plans provided by GPT-4 and the coaches were very close, with no significant differences. These findings suggest that, according to expert evaluations, GPT-4 is capable of generating fitness plans that are nearly on par with those created by professional fitness coaches.

4.1 Performance of GPT-4 Generated Plans

In our evaluation, the primary advantage of GPT-4 lies in its robust personalization capabilities, attributable to its deep learning algorithms and extensive data processing capacity (X. Wang et al., 2022). However, the performance of GPT-4 in meeting personalized needs did not fully achieve statistical significance, which could be attributed to the small sample size and limitations during model training (Kojima, Gu, Reid, Matsuo, & Iwasawa, 2022). Additionally, there were no significant differences found in effectiveness, comprehensiveness, and safety when compared to human coaches, indicating that GPT-4 can provide fitness guidance of a quality comparable to that of professional fitness coaches in these critical dimensions. This study further supports the findings of

Washif et al., who noted that Chat GPT can quickly generate detailed resistance training prescriptions (Washif et al., 2024). However, Wang et al. (2023) pointed out that the application of Chat GPT in the field of exercise science is not entirely satisfactory. The divergence in these viewpoints may be due to several factors. First, Chat GPT is based on a large language model (LLM) driven by the GPT architecture, which processes natural language through neural networks and can be trained on large-scale multilingual text data to generate human-like responses (Kojima et al., 2022). The rapid pace of updates and iterations means that the language models used in different studies may vary. Second, the prompts used in the studies differ. Properly crafted prompts can significantly enhance the response quality of large language models (LLMs) (Xu et al., 2023). Fulford and Ng (2023) have launched an online course which suggests some basic guidelines for developing prompts (A. J. D. A. Ng, 2023). Akin (2023) confirmed the efficacy of these prompt resources (Li, Hammoud, Itani, Khizbullin, & Ghanem, 2023).

4.2 Accessibility and Convenience of GPT-4 Reduces the Cost of Weight Loss

In China, a developing country, the vast demand for high-quality fitness guidance is mismatched by the limited public sports and fitness services, making it difficult for the general population to access personalized fitness guidance (Ren et al., 2021). A survey of selected fitness centers revealed that personal training sessions primarily focused on weight loss and body shaping cost approximately 240 yuan (about $37 USD) per session. This represents a significant expense, particularly in cities with relatively limited sports resources. Consequently, the widespread implementation of effective and personalized health interventions across society continues to face numerous challenges (Mauro, Taylor, Wharton, & Sharma, 2008). However, the advent of artificial intelligence (AI) technology introduces new strategies in the health promotion realm (Lee, Bubeck, & Petro, 2023). GPT-4, demonstrating exceptional performance across various domains, emerges as a preferred option due to its convenience and accessibility. The training plans and dietary recommendations generated by GPT-4, based on users' detailed information, can rival the guidance provided by professional fitness coaches, offering an efficient and cost-effective approach to fitness. The advantages of using AI models like GPT-4 not only lower the cost of personalized fitness guidance but also significantly enhance its accessibility, especially in areas with limited fitness resources. This AI-driven approach aligns with the global trend of leveraging technology to overcome barriers in medical and health services (Zakerabasali, Ayyoubzadeh, Baniasadi, Yazdani, & Abhari, 2021). Future research should focus on assessing the efficacy of GPT-4 in delivering personalized health and fitness interventions that seamlessly integrate into people's daily lives. This not only addresses the accessibility issues of fitness guidance but also plays a significant role in combating obesity and promoting healthier lifestyles worldwide.

5 Limitations

This study has several key limitations. Firstly, the limited sample representation, consisting of only one participant with a specific background, impacts the external validity of the research. Although Chat GPT's responses showed some repetitiveness (Hadi et al., 2023). future studies should aim to expand and diversify the participant pool. Secondly, the 16-week evaluation period may be insufficient to comprehensively reflect the long-term effects of fitness plans (Douketis, Macie, Thabane, & Williamson, 2005). and the research's applicability is constrained by its focus on short-term outcomes and lack of detailed implementation guidelines. Future research should extend the evaluation period, incorporate more objective metrics, and refine the personalization of fitness plans. Additionally, GPT-4's recommendations may be influenced by limitations in its training data (such as physical fitness test results and body composition analysis), and it has inherent constraints, such as the inability to provide real-time correction of exercise techniques or psychological support (LeCun, Bengio, & Hinton, 2015). Nevertheless, the application of GPT-4 represents a significant advancement in bridging the gap between artificial intelligence and human expertise. Future research should focus on optimizing the integration of AI and human professional knowledge to achieve greater breakthroughs in personalized fitness guidance.

6 Conclusion

This study evaluated the potential application of GPT-4 as a virtual fitness coach in designing personalized weight loss and fitness plans. While GPT-4 demonstrated relative advantages in personalized design, it did not show significant differences in effectiveness, comprehensiveness, and safety compared to human fitness coaches. These findings suggest that GPT-4 can contribute valuable insights to fitness guidance. Therefore, future research should explore the integration of GPT-4 with the expertise of human coaches to overcome technological limitations and enhance the practical application of personalized fitness plans. This study provides a preliminary theoretical and practical foundation for the future development of artificial intelligence in the field of personalized health promotion.

CRediT authorship contribution statement

GCL, HSL, YQS, YL and GDZ designed this study. GCL and HSL jointly collected training programs. YQS and HSL analyzed data. GCL, YQS and HSL wrote the first draft of the manuscript. All authors contributed to the final manuscript.

Role of the funding source

This work was supported by the Horizontal Scientific Research Project of Southwest University in the year of 2023 (No. 2308017).

Declaration of competing interest

None.

Acknowledgment

We would like to thank all of the researchers who kindly provided us with the data necessary to complete this study.

Reference

Almeida, L. G., Dera, A., Murphy, J., & Santosa, S. J. C. o. (2024). Improvements in cardiorespiratory fitness, muscle strength and body composition to modest weight loss are similar in those with adult-versus childhood-onset obesity. *14*(1), e12623.

Arslan, S. J. A. o. b. e. (2023). Exploring the potential of Chat GPT in personalized obesity treatment. *51*(9), 1887-1888.

Bhattarai, K., Oh, I. Y., Sierra, J. M., Payne, P. R., Abrams, Z. B., & Lai, A. M. J. B. t. P. S. f. B. (2023). Leveraging GPT-4 for Identifying Clinical Phenotypes in Electronic Health Records: A Performance Comparison between GPT-4, GPT-3.5-turbo and spaCy's Rule-based & Machine Learning-based methods.

Blüher, M. J. T. L. (2008). Efficacy and safety of the weight-loss drug rimonabant. *371*(9612), 555-556.

Bompa, T. O., & Buzzichelli, C. (2019). *Periodization-: theory and methodology of training*: Human kinetics.

Bouchard, C., Depres, J. P., & Tremblay, A. J. O. r. (1993). Exercise and obesity. *1*(2), 133-147.

Buford, T. W., Roberts, M. D., & Church, T. S. J. S. m. (2013). Toward exercise as personalized medicine. *43*, 157-165.

Church, T. S., Martin, C. K., Thompson, A. M., Earnest, C. P., Mikus, C. R., & Blair, S. N. J. P. o. (2009). Changes in weight, waist circumference and compensatory responses with different doses of exercise among sedentary, overweight postmenopausal women. *4*(2), e4515.

Crookham, J. J. P. C. C. i. O. P. (2013). A guide to exercise prescription. *40*(4), 801-820.

Deng, S. (2015). Exercise Physiology, 3rd Edition. *Higher Education Press*, *330*, 330.

Derickson, M., Phillips, C., Barron, M., Kuckelman, J., Martin, M., & DeBarros, M. J. T. A. J. o. S. (2018). Panniculectomy after bariatric surgical weight loss: analysis of complications and modifiable risk factors. *215*(5), 887-890.

Douketis, J., Macie, C., Thabane, L., & Williamson, D. J. I. j. o. o. (2005). Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice. *29*(10), 1153-1167.

Fanning, J., Mullen, S. P., & McAuley, E. J. J. o. m. I. r. (2012). Increasing physical activity with mobile devices: a meta-analysis. *14*(6), e2171.

Galiuto, L., Fedele, E., Vitale, E., & Lucini, D. J. C. S. M. R. (2019). Personalized exercise prescription for heart patients. *18*(11), 380-381.

Geerling, W., Mateer, G. D., Wooten, J., & Damodaran, N. J. A. a. S. (2023). Is ChatGPT smarter than a student in principles of economics. *4356034*.

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., . . . Mirjalili, S. J. A. P. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage.

Iqbal, R. K., Masood, F., & Ikram, S. J. N. J. o. H. S. (2019). How Obesity Affects Our Health. *4*(3), 113-118.

Issurin, V. B. J. S. m. (2010). New horizons for the methodology and physiology of training periodization. *40*, 189-206.

James, W. P. T. J. C. c. (2018). Obesity: a global public health challenge. *64*(1), 24-29.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. J. A. i. n. i. p. s. (2022). Large language models are zero-shot reasoners. *35*, 22199-22213.

LeCun, Y., Bengio, Y., & Hinton, G. J. n. (2015). Deep learning. *521*(7553), 436-444.

Lee, P., Bubeck, S., & Petro, J. J. T. N. E. j. o. m. (2023). Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. Reply. *388*(25), 2400-2400.

Li, G., Hammoud, H., Itani, H., Khizbullin, D., & Ghanem, B. J. A. i. N. I. P. S. (2023). Camel: Communicative agents for" mind" exploration of large language model society. *36*, 51991-52008.

Lim, G. B. J. N. R. C. (2024). Weight loss from surgery or drug therapy reduces blood pressure. *21*(4), 218-218.

Locke, E. A. (1987). Social foundations of thought and action: A social-cognitive view. In: Academy of Management Briarcliff Manor, NY 10510.

Mauro, M., Taylor, V., Wharton, S., & Sharma, A. M. J. E. j. o. i. m. (2008). Barriers to obesity treatment. *19*(3), 173-180.

Mazumdar, H., Chakraborty, C., Sathvik, M., Panigrahi, P. K. J. I. J. o. B., & Informatics, H. (2023). GPTFX: A Novel GPT-3 based framework for mental health detection and explanations.

McInnis, K. J. J. C. a. d. (2000). Exercise and obesity. *11*(2), 111-116.

Ng, A. J. D. A. (2023). ChatGPT prompt engineering for developers.

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., . . . Abera, S. F. J. T. l. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013. *384*(9945), 766-781.

Petridou, A., Siopi, A., & Mougios, V. J. M. (2019). Exercise in the management of obesity. *92*, 163-169.

Ren, P., Liu, Z. J. I. J. o. E. R., & Health, P. (2021). Efficiency evaluation of China's public sports services: a three-stage dea model. *18*(20), 10597.

Sabin, J. A., Marini, M., & Nosek, B. A. J. P. o. (2012). Implicit and explicit anti-fat bias among a large sample of medical doctors by BMI, race/ethnicity and gender. *7*(11), e48448.

Schwartz, M. B., Chambliss, H. O. N., Brownell, K. D., Blair, S. N., & Billington, C. J. O. r. (2003). Weight bias among health professionals specializing in obesity. *11*(9), 1033-1039.

Shi, Q., Wang, Y., Hao, Q., Vandvik, P. O., Guyatt, G., Li, J., . . . Ge, L. J. T. L. (2024). Pharmacotherapy for adults with overweight and obesity: a systematic review and network meta-analysis of randomised controlled trials. *403*(10434), e21-e31.

Slomski, A. J. J. (2022). Weight Loss Is Still Substantial a Decade After Bariatric Surgery. *328*(5), 415-415.

Štempeľová, I., Takáč, O., Hudáková, H. J. J. o. A. N. S., & Researches, E. (2023). Obesity as a 21st century pandemic. *7*(6), 449-454.

Szabo, A. (2023). ChatGPT a breakthrough in science and education: Can it fail a test?

Villareal, D. T., Chode, S., Parimi, N., Sinacore, D. R., Hilton, T., Armamento-Villareal, R., . . . Shah, K. J. N. E. J. o. M. (2011). Weight loss, exercise, or both and physical function in obese older adults. *364*(13), 1218-1229.

Vogel, T., Brechat, P. H., Leprêtre, P. M., Kaltenbach, G., Berthel, M., & Lonsdorfer, J. J. I. j. o. c. p. (2009). Health benefits of physical activity in older patients: a review. *63*(2), 303-320.

Wang, M., Wang, M., Xu, X., Yang, L., Cai, D., & Yin, M. J. I. T. o. L. T. (2023). Unleashing ChatGPT's power: A case study on optimizing information retrieval in flipped classrooms via prompt engineering.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., . . . Zhou, D. J. a. p. a. (2022). Self-consistency improves chain of thought reasoning in language models.

Washif, J., Pagaduan, J., James, C., Dergaa, I., & Beaven, C. J. B. o. s. (2024). Artificial intelligence in sport: Exploring the potential of using ChatGPT in resistance training prescription. *41*(2), 209-220.

Xu, B., Yang, A., Lin, J., Wang, Q., Zhou, C., Zhang, Y., & Mao, Z. J. a. p. a. (2023). Expertprompting: Instructing large language models to be distinguished experts.

Zakerabasali, S., Ayyoubzadeh, S. M., Baniasadi, T., Yazdani, A., & Abhari, S. J. H. I. R. (2021). Mobile health technology and healthcare providers: systemic barriers to adoption. *27*(4), 267-278.

Zhi, Y. (2022). Diet control and mineral intake help obese students lose weight and prevent some diseases. *International Conference on Biotechnology, Life Science*

Zhu, W., Geng, W., Huang, L., Qin, X., Chen, Z., Yan, H. J. J. o. S., & Science, H. (2024). Who could and should give exercise prescription: Physicians, exercise and health scientists, fitness trainers, or ChatGPT? , *13*(3), 368.