©Journal of Sports Science and Medicine (2026) **25,** 235-261
http://www.jssm.org DOI: https://doi.org/10.52082/jssm.2026.235

Research article

ChatGPT Outperforms Personal Trainers in Answering Common Exercise Training Questions

Brecht D'hoe [1]✉, Daniel Kirk [2], Jan Boone [3] and Alessandro Colosio [4]
[1] Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium; [2] Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas Hospital, Westminster Bridge Road, London SE1 7EH, UK; [3] Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium; [4] Inter-University Laboratory of Human Movement Biology, Université Saint-Etienne, Saint-Etienne, France

Abstract
Since its launch, the chatbot ChatGPT has gained significant popularity and may serve as a valuable resource for evidence-based exercise training advice. However, its capability to provide accurate and actionable exercise training information has not been systematically evaluated. This study assessed ChatGPT's proficiency by comparing its responses to those of human personal trainers. Nine currently active level 4 (European Qualification Framework (EQF)) personal trainers (PTs) submitted their most frequently asked exercise training questions along with their own answers to them, and these questions were then posed to ChatGPT (version 3.5). Responses from both sources were evaluated by 18 PTs and 9 topic experts, who rated them on scientific correctness, actionability, and comprehensibility. Scores for each criterion were averaged into an overall score, and group means were compared using permutation tests. ChatGPT outperformed PTs in six of nine questions overall, with higher ratings in scientific correctness (5/9), comprehensibility (6/9), and actionability (5/9). In contrast, none of the responses from PTs were higher than those from ChatGPT for any question or metric. Our results suggest that ChatGPT can be used as a tool to answer questions that are frequently asked to PTs, and that chatbots may be useful for delivering informational support relating to physical exercise.

Key words: Artificial Intelligence, exercise, natural language processing, machine learning, training guidance.

Introduction

In recent years, advancements in digital innovation have made cutting-edge technologies a part of daily life, largely driven by the widespread availability and ease of use offered by the Internet, smartphones, and other electronic devices. One of the most pronounced examples of this has been the introduction of highly sophisticated chatbots that utilize natural language processing (large language models, LLMs), a subset of artificial intelligence, to interpret and reply to text input, simulating human conversation. Of the various options available, OpenAI's ChatGPT has been the most prominent, with around 200 million monthly active users at the time of writing (Malik, 2023; Moore, 2023). ChatGPT's versatility, enhanced context understanding, and advanced natural language processing capabilities enable users to quickly obtain foundational knowledge in a particular field more efficiently than browsing through search engine results (Ray, 2023).

The domain of exercise has great potential for the implementation of chatbots due to the number of people who tend to seek out information related to this field online (Lupton, 2020; Digital Health: Critical and Cross-Disciplinary Perspectives). While this has traditionally occurred via individual websites accessed via search engines, chatbots offer distinct advantages. For example, chatbots can aggregate data from multiple sources, providing a more comprehensive and balanced view in one convenient location rather than requiring users to visit several different web pages. Furthermore, the possibilities to engage in conversation allow users to ask follow-up questions and seek clarification, enhancing the learning experience in ways that traditional web platforms often cannot. Within the domain of exercise, chatbots can also have practical advantages over alternative and more conventional ways of obtaining information, such as communicating with fitness professionals. For example, chatbots are not restricted to office hours and are therefore constantly available to answer fitness-related questions. They can also generate answers instantly, and many are free to use.

The promising role of chatbots has been recently explored in the domains of public health, medicine, and nutrition and dietetics (Gardiner et al., 2017; Gabarron et al., 2020; Kirk et al., 2023; Morreel et al., 2023). However, research on the use of chatbots in the domain of exercise and training primarily focuses on their role in promoting physical activity and prescribing exercise. For example, Piao et al. found that a chatbot served as a cost-effective tool to enhance conscious decision-making, encouraging increased physical activity in the workplace (Piao et al., 2020). Zhu et al. found that ChatGPT combines the perception of the physician, exercise scientist, and personal trainer to successfully create exercise plans (Zhu et al., 2024). Conversely, Düking et al. (2024) found that running experts rated training plans created by ChatGPT as suboptimal; however, they concluded that this was also dependent on the depth of information provided in the prompt input (Düking et al., 2024). For both clinical and healthy resistance-trained populations, it has been shown that ChatGPT can provide suitable exercise prescriptions, though findings regarding its safety were mixed (Dergaa et al., 2024; Washif et al., 2024; Zaleski et al., 2024).

While ChatGPT shows promise in supporting the field of exercise prescription, it may lack personalization and require further modification before implementation (Dergaa et al., 2024; Washif et al., 2024; Zaleski et al., 2024), with the provided information being rather accurate but not comprehensive (Zaleski et al., 2024). Additionally,

Received: 09 August 2025 / Accepted: 18 December 2025 / Published (online): 01 February 2026

studies that have investigated the role of chatbots in promoting physical activity lack the consistent usage of measurements and reporting of outcome evaluations (Oh et al., 2021). Studies investigating ChatGPT in exercise prescription may have a biased scoring process, as the reviewers evaluating ChatGPT's answers do not appear to have been blinded (Dergaa et al., 2024; Düking et al., 2024). It has also been noted that there is insufficient evidence for the acceptability and practical feasibility of chatbots in general (Han et al., 2023). Hence, while there is promise, chatbots in exercise science remain understudied, with only a limited number of domains assessed and no direct comparison with human alternatives. Therefore, the objective of the current study was to compare answers from ChatGPT to frequently asked exercise training questions with those of human personal trainers (PTs). We hypothesized that ChatGPT would perform better than human PTs in answering frequently asked exercise training questions. By investigating this aim, we provide scientific support for the use of chatbots as a source of information in the exercise field, which may reduce the workload burden for PTs and support trainees seeking knowledge.

Methods

Experimental Protocol

The study took place from March to July 2024. The chronological flow of the experimental protocol can be found in Figure 1. PTs were asked via Microsoft Forms to provide their most frequently asked questions related to exercise training, as well as their own answers to them. The PTs were requested to answer these questions in the same fashion that they would for a potential client. PTs could provide the question and answer in Dutch, French, and English, depending on their linguistic competence and preference.

Exclusion criteria for the questions included failure to refer to exercise training or the implementation of lifestyle and health-related factors; answers that were not articulated in a manner suitable for direct communication with a potential client (e.g., "it is a boring question", "it depends"); and those that fell below the minimum word count of 80 or exceeded the maximum of 300 words. None of the PTs were made aware of the goal of the study or the involvement of ChatGPT prior to their contribution to prevent influencing the questions or answers they provided. Following submission of their questions and answers, they were informed about the goal of the study and were given the opportunity to retract their input if they wished to; however, none of the participants chose to do so.

After applying the exclusion criteria, the remaining questions formed the final set of nine questions and answers from nine different personal trainers used in the study (Table 1). While this falls slightly short of our desired sample size (N = 17, based on an effect size of a mean difference, a standard deviation of 1 between the groups, and for each question, a desired power of 0.8 and α of 0.95 (two-tailed), we initiated analysis after recruiting 9 participants due to a low response rate and time limitations. While this reduces the power of our study slightly, sample sizes of a similar number have been reported elsewhere in peer-reviewed, published literature (Kirk et al., 2023). These questions were then asked to ChatGPT (Version 3.5) in their original form. ChatGPT's customization settings were left in the default settings, and no additional information other than the question was provided. The questions were only asked once, and the answer was recorded.

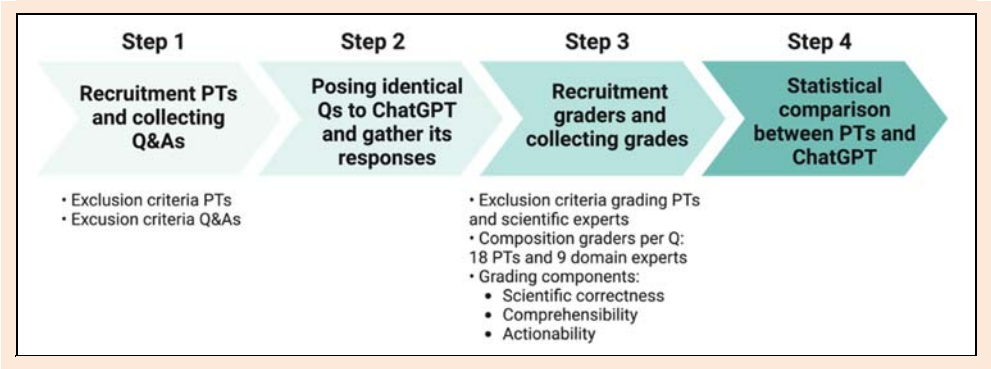


Figure 1. Chronological flow chart of the experimental protocol used in the current study.

Table 1. The most frequently asked questions reported by the PTs categorized by theme.

	Question	Question Topic
1	How often do I need to train for the training to be effective?	Training frequency
2	How do I lose abdominal fat?	Training for fat loss
3	How can I maintain my motivation to work out?	Training motivation
4	How do I train to lose weight?	Training for fat loss
5	Is cardio better to lose weight than strength workouts?	Training for fat loss
6	How can I train around pain?	Training around pain
7	Working out at what time of the day leads to most effective results?	Timing of the day for training
8	Which exercises are best for losing weight?	Training for fat loss
9	How do I get a tighter stomach?	Training for fat loss

The answers to each of the questions were graded by other PTs and scientific experts in the field of the questions. Each set of answers was graded by a total of 27 graders. The order of answers to the questions was randomized in the grading document to reduce the chances of pattern recognition and unblinding of the graders. One of the three scoring components was scientific correctness, and therefore, nine of the 27 graders (1/3) were scientific domain experts in the field of the question. The other 18 graders were personal trainers who graded all nine questions, irrespective of the topic. As with the PTs providing the question and answers, graders were initially blinded to the actual goal of the study and thus the involvement of ChatGPT. After receiving the evaluation of the answers, graders were informed about the goal of the study and were provided the opportunity to subtract their contribution if they disagreed with the procedure; however, none of the graders chose to do so.

The grading was based on three components deemed relevant for determining the quality of an answer to a hypothetical knowledge-seeker asking a question: "Scientific correctness", reflecting how accurately an answer reflects the current state of knowledge in the scientific domain to which the question belongs; "Comprehensibility", capturing how well the answer could be expected to be understood by the layman receiving the answers, and; "Actionability", the extent to which the answer to the question contains information that is useful and can be acted upon by the hypothetical layman asking the question. The rubric describing how each component should be scored that was sent to the graders is presented in Table 2. Each component could be scored between 0 and 10, with 0 representing a complete failure to satisfy a given component in the answer and 10 reflecting a perfect score on the given component. The overall score was acquired by averaging the scores of all three components.

Participants – contributors

Personal trainers (PTs) providing questions and answers were recruited through relations within Ghent University and the European Register of Exercise Professionals (EREPS). Contributing PTs were required to hold at least a European Qualification Framework (EQF) Level 4 certification or an equivalent or higher qualification. This criterion ensured inclusion of professionally qualified PTs while excluding self-proclaimed ones, and participants were further required to be actively engaged in personal training. All provided online written informed consent to participate in the study.

Participants – graders

To grade the answers provided by the contributing PTs, nine scientific experts per question topic were recruited from universities or research institutions and were required to comply with the inclusion criteria: minimum academic level of having attained a PhD, being currently affiliated with a research institute (either academic or private) and having expertise in the topic of the question based on topic-related peer-reviewed scientific article output. Hence, the scientific experts (1/3 of all graders) only rated the answers to the question relating to their domain of expertise across all three components (comprehensibility, actionability, and scientific correctness). In alignment with their professional responsibilities relating to relaying information to trainees and thus representing comprehensibility and actionability, the remaining 18 graders (2/3 of all graders) were personal trainers. Accordingly, the grading PTs rated all answers to all questions across the three components. Grading PTs were recruited through contacts of Ghent University and the Belgian commercial gym franchise 'Jims'. The same inclusion criteria applied to the grading PTs as for those contributing to questions and answers, implying having obtained a minimum of EQF level 4 or equivalent or higher, and being actively engaged in personal training.

Ethical approval

The study was conducted according to the Declaration of Helsinki, and all procedures were approved by the ethics committee of the faculty of psychology and educational sciences, Ghent University (reference code: 2024-005). After receiving a description of the study process, all participants provided online written informed consent to participate in the current study and could withdraw at any time (including after the actual goal of the study and the involvement of ChatGPT was made clear). All participants were informed that their participation would be made pseudonymous.

Statistical analysis

Differences between the grades of the answers overall and for each grading component were determined using permutation tests with the function perm.test from the package

Table 2. The explanations of each of the grading criteria sent to the graders.

Grading Component	Description
Scientific correctness	How accurately each answer reflects the current state of knowledge in the scientific domain to which the question belongs. The requested word count of the answers (100-300 words) and the natural limitations on detailed explanation and nuance this imposes should be kept in mind when grading scientific correctness. The target audience of the layman and their expected level of scientific knowledge and training understanding should also be kept in mind.
Comprehensibility	How well the answer could be expected to be understood by the layman. Comprehensibility should pertain mostly to the content of the answer, though if grammatical errors hinder comprehensibility, then this may also be considered.
Actionability	The degree to which the answers to the questions contain information that is useful and can be acted upon by the hypothetical layman asking the question. For example, whilst bariatric surgery may represent an effective weight loss strategy for morbidly obese individuals, this would not be a helpful suggestion for someone with a BMI of 27 looking to lose a little weight. Hence, such an answer would score poorly on this component.

jmuOutlier in R software (R: The R Project for Statistical Computing). Permutation tests were used to test for group differences with the test statistic set to the mean. We further tested the robustness of our results with a sensitivity analysis using the median as the test statistic to account for the potential effects of any outliers. For both the overall score and the scores of the components for each question, p-values were approximated from 100,000 simulations to gauge the strength of evidence of a difference between the groups. Additional analysis on the comprehensibility of the answers was done using the Flesch Reading Ease (FRE) score and the Flesch–Kincaid Grade Level from the textstat library (version 0.7.10) in Python (version 3.11.7) using the flesch_reading_ease and flesch_reading_ease functions, respectively. These scores were calculated for each answer in both groups and compared using two-sided permutation tests using the permutation_test (N permutation = 9999, permutation type = 'independent') from the scipy.stats library (version 1.11.4). To evaluate the consistency among graders, inter-rater reliability (IRR) was assessed using the intraclass correlation coefficient (ICC) function from the irr package in R software. ICC values could be calculated for all grading PTs and for the scientific expert graders within the topic "training for fat loss", as these were the only instances with multiple raters per item - a prerequisite for ICC computation. Cut-off values for the interpretation of the ICC values were based on Koo and Li (Koo and Li, 2016). Furthermore, the level of agreement in question grading was calculated across all scientific expert graders to assess their consistency, using descriptive statistics (standard deviation) for each answer set (defined by topic, question, criterion, and source: ChatGPT or PT).

Results

Study sample

After applying the exclusion criteria, nine questions and answers from nine PTs were eligible for inclusion out of the initial 47 inputs received. The age of the nine PTs providing the questions and answers ranged from 19 to 48 (mean age = 38 years). Their experience in personal training varied from 0.5 to 15 years, with a mean of 5.3 years. The nationalities of the PTs were Spanish, Swedish, Malaysian, Belgian, French, Danish, Pakistani, and two were Dutch. All of them had obtained a European Qualification Framework 4 (EQF4) personal training certificate and were all currently active in personal training. Additional personal training-related qualifications that the PTs held are displayed in Table S1.

Of the nine included questions, five were related to training for fat/weight loss, one to training frequency, one to training motivation, one to training around pain, and one to timing of the day for training. The answers from the personal trainers across the nine questions ranged from 80 to 280 words, with an average of 140 words, whereas those from ChatGPT ranged from 181 to 274 and had an average of 241 words.

The scientific experts who graded the topic-specific answers represented a diverse range of countries of employment. Among the nine reviewers assigned to each topic, at least five different countries were represented, with one topic including experts from up to eight countries. Furthermore, for three of the topics, seven out of nine reviewers were active professors in the respective field, while the remaining two topics included no fewer than five and six professors among the reviewers, respectively (Table S2). As for the grading PTs, five different nationalities were represented, with the majority being Belgian. All grading PTs held an EREPS level 4 personal training qualification or an equivalent—or higher—certification (Table S3).

Grades of answers from PTs and ChatGPT

Figure 2 illustrates the overall scores for the answers to each question, and the key summary statistics for the overall scores of the answers to each question can be found in Table S4. There was strong evidence that the mean overall scores for ChatGPT were significantly higher for six of the questions, specifically, Question 1 (7.57 for ChatGPT vs 6.69 for PTs; $p = 0.0178$), Question 2 (7.96 for ChatGPT vs 5.49 for PTs; $p = < 0.0001$), Question 3 (7.98 for ChatGPT vs 7.01 for PTs; $p = 0.0004$), Question 6 (7.99 for ChatGPT vs 7.07 for PTs; $p = 0.0071$), Question 7 (8.15 for ChatGPT vs 6.36 for PTs; $p = 0.0001$), Question 8 (7.73 for ChatGPT vs 6.65 for PTs; $p = 0.0047$). The scores of ChatGPT also tended to be higher for Question 9, although the strength of evidence was weaker ($p = 0.0535$). The mean overall scores were similar for Questions 4 and 5. The answers to each of the questions from the PTs and ChatGPT can be seen in Table S5, and the grades for them in Table S6.

Figure 3, Figure 4 and Figure 5 show boxplots of the grades for the individual components and tables of their summary statistics can be found in Tables S7-S9. The mean scores for scientific correctness were higher for ChatGPT for five of the questions, specifically, Question 1 (8.26 for ChatGPT vs 5.43 for PTs; $p < 0.0001$), Question 2 (8.04 for ChatGPT vs 5.22 for PTs; $p < 0.0001$), Question 3 (7.26 for ChatGPT vs 6.33 for PTs; $p = 0.0036$), Question 7 (7.89 for ChatGPT vs 5.93 for PTs; $p = 0.0002$) and, Question 8 (7.59 for ChatGPT vs 6.19 for PTs; $p = 0.026$). No differences were observed for the other four questions (Figure 3).

There was strong evidence for higher mean scores of comprehensibility for ChatGPT in six instances, specifically, Question 2 (8.63 for ChatGPT vs 6.15 for PTs; $p < 0.0001$), Question 3 (8.48 for ChatGPT vs 7.37 for PTs; $p = 0.0077$), Question 6 (8.37 for ChatGPT vs 7.11 for PTs; $p = 0.0011$), Question 7 (8.52 for ChatGPT vs 6.70 for PTs; $p = 0.0001$), Question 8 (8.26 for ChatGPT vs 7.26 for PTs; $p = 0.0039$) and, Question 9 (8.04 for ChatGPT vs 7.30 for PTs; $p = 0.0438$). No differences were observed for the remaining Questions 1, 4, and 5 (Figure 4).

Finally, there was strong evidence for higher mean actionability for ChatGPT on five occasions, specifically, Question 2 (7.22 for ChatGPT vs 5.11 for PTs; $p < 0.0001$), Question 3 (8.20 for ChatGPT vs 7.33 for PTs; $p = 0.0213$), Question 6 (7.78 for ChatGPT vs 6.48 for PTs; $p = 0.0127$), Question 7 (8.04 for ChatGPT vs 6.44 for PTs; $p = 0.0023$), Question 8 (7.33 for ChatGPT vs 6.52 for PTs; $p = 0.0355$), with no clear differences for the remaining Questions 1,4, 5 and 9 (Figure 5). A visual representation of the questions that score higher for ChatGPT, both overall and for each grading component can be seen in Figure 6.

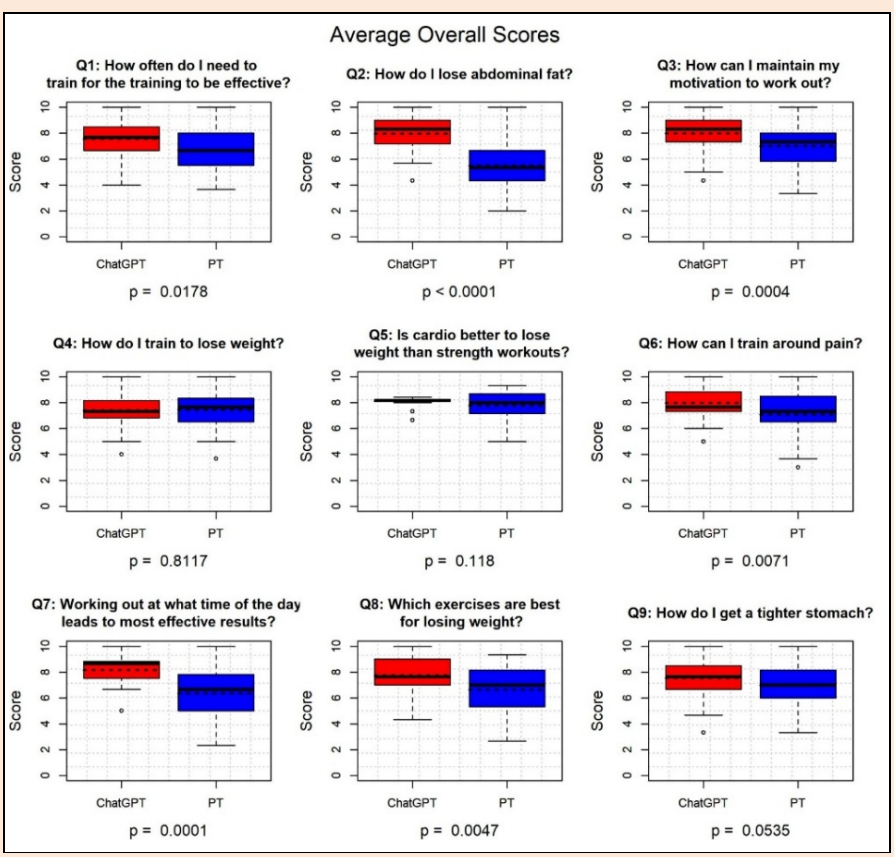


Figure 2. **Overall grades for the answers to the questions given by ChatGPT (green, left) and PTs (blue, right). The upper and lower bounds of the boxplots present the interquartile range, the red dotted line inside the box represents the average, and the black solid line represents the median of the data. The upper and lower lines beyond the box (whiskers) are the highest and lowest values, respectively, excluding outliers.**

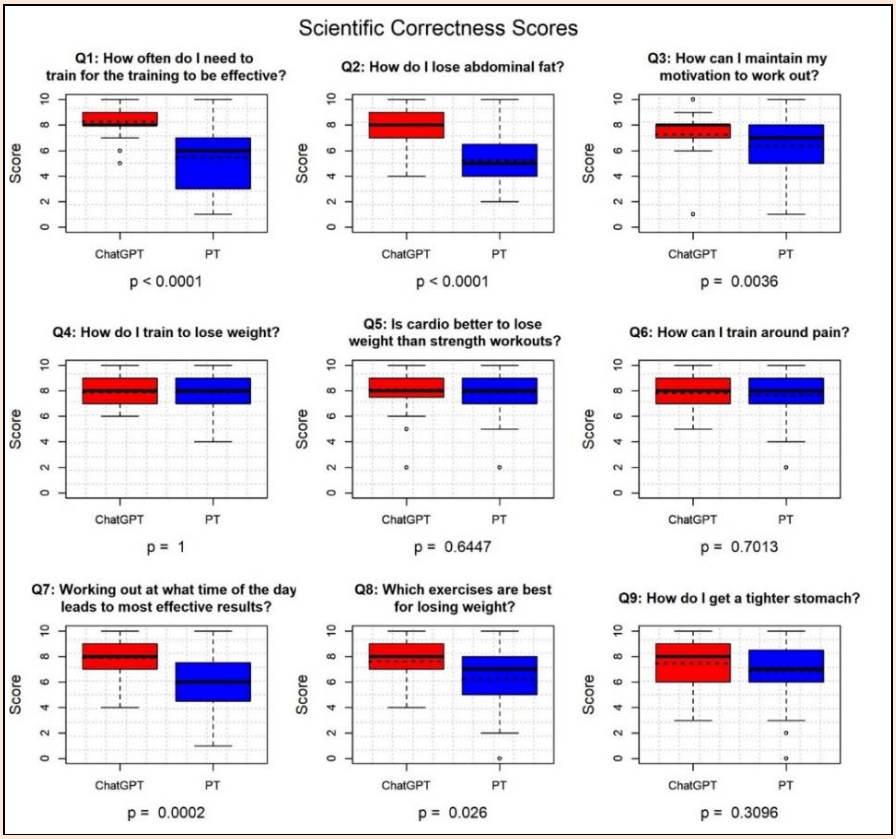


Figure 3. **Grades for the component "Scientific correctness" for each question for ChatGPT and the PTs.**

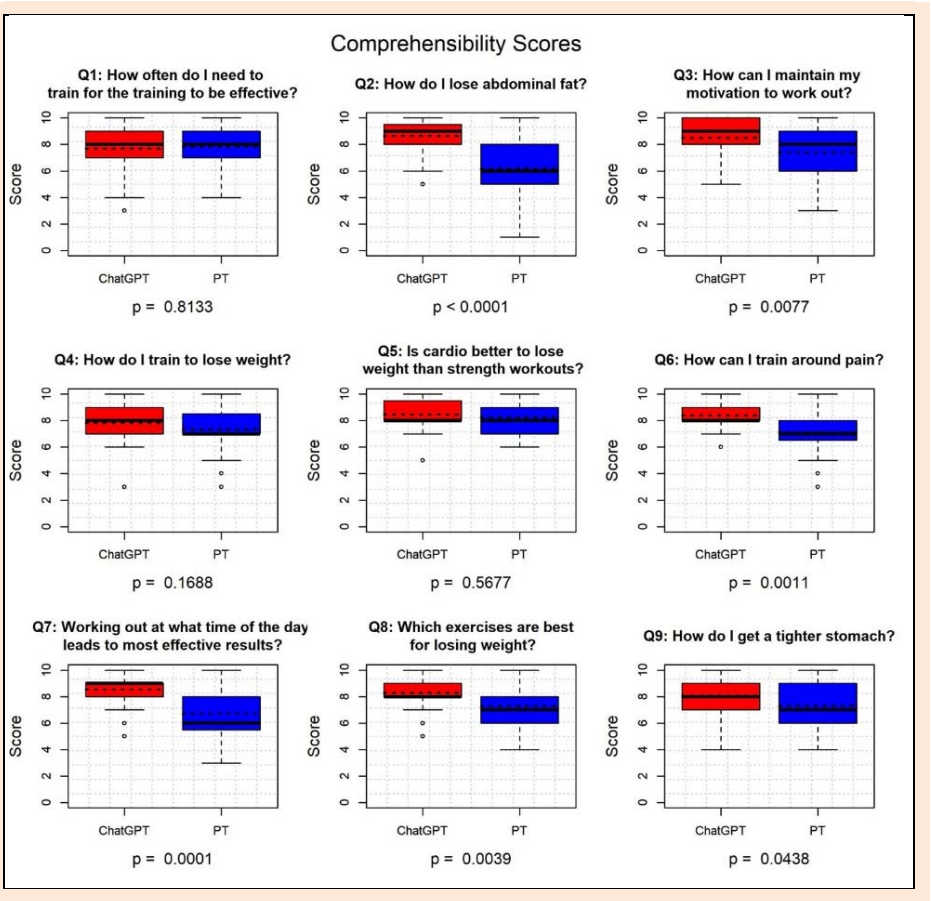


Figure 4. Grades for the component "Comprehensibility" for each question for ChatGPT and the PTs.

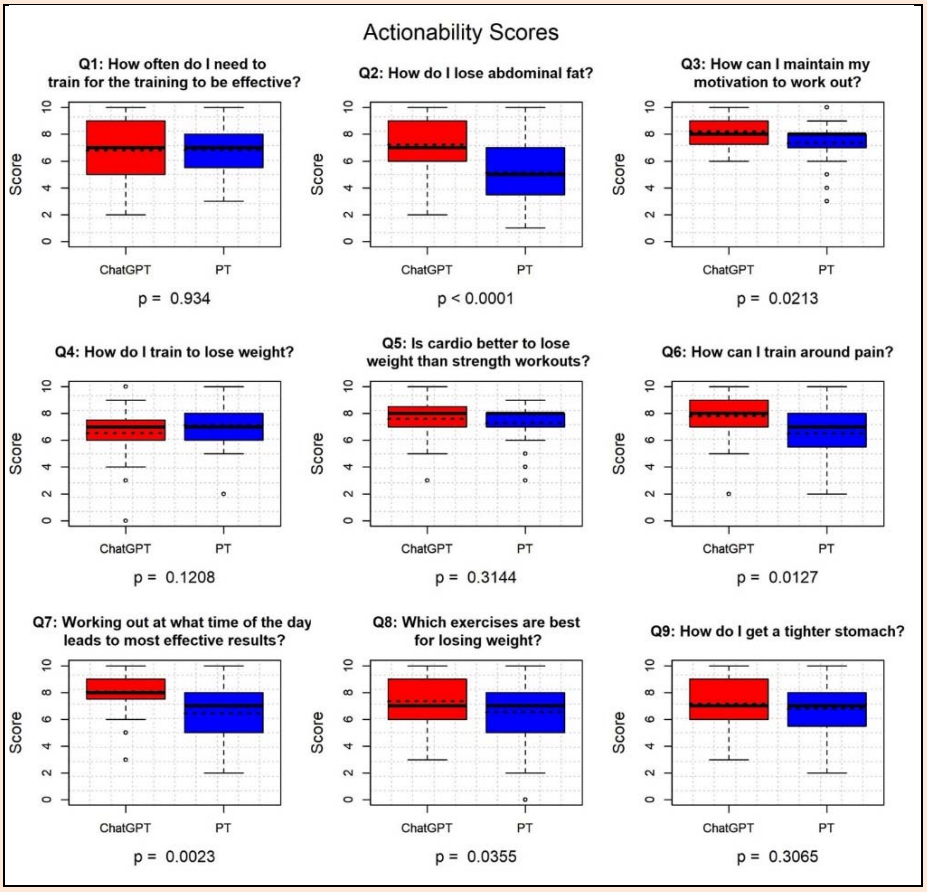


Figure 5. Grades for the component "Actionability" for each question for ChatGPT and the PTs.

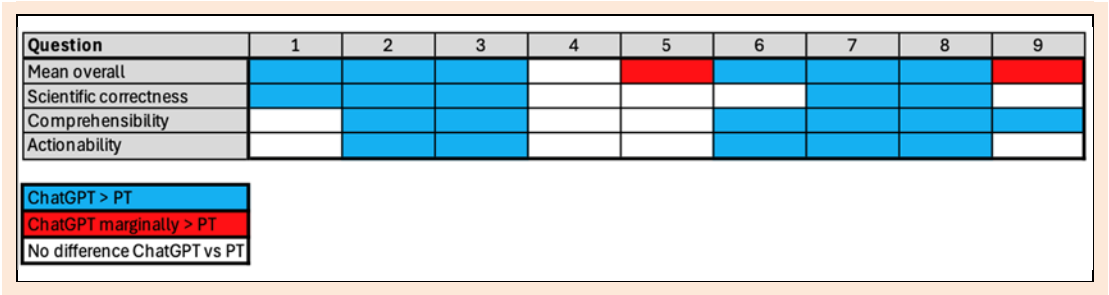


Figure 6. The grading components outlined for each question with a color scheme representing occasions in which ChatGPT scores higher (blue) or marginally higher (red) than PTs.

Sensitivity analysis

Sensitivity analysis using the median instead of the mean as the test statistic for group differences was also conducted. These results can be seen in Table S10. In general, the *p*-values of the permutations test tended to be higher when the median was used as the test statistic. This could suggest that the central tendency appeared to be more similar between the groups, with scores closer to the limits of the response (i.e., 0 or 10) driving mean differences. Nevertheless, although the median differences were less pronounced, there was still clear evidence for higher scores for answers from ChatGPT compared to those from PTs (Figure 6). We also analyzed comprehensibility by comparing Flesch Reading Ease (FRE) scores and the Flesch–Kincaid Grade Levels between the groups, though we identified no significant differences here (Table S11 and Table S12). In addition, inter-rater reliability among the grading PTs, who evaluated all answers to all nine questions, showed good to excellent intraclass correlation coefficients (ICC) values overall, except for the ChatGPT answer set under the criterion scientific correctness, which demonstrated lower inter-rater reliability (Table S13). The ICCs among the scientific expert graders evaluating the answers to the five training for fat loss questions were generally higher for the PT-provided answers than for those generated by ChatGPT across all three grading criteria (Table S14). The per-question agreement analysis among all scientific expert graders indicated generally higher grading consistency for ChatGPT compared to PTs, with the greatest variability observed for the criterion of scientific correctness (Table S15 and Table S16).

Discussion

ChatGPT outperforms human PTs in answering common exercise training questions

The results of the current study provide compelling evidence supporting our hypothesis that responses from ChatGPT outperform human PTs when answering common exercise training questions, as indicated by higher scores in six of the nine questions for the mean overall score. In contrast, overall scores for human PTs failed to exceed those of ChatGPT in any of the questions. A similar pattern was also seen for each of the individual grading criteria from which the overall scores were composed (i.e., scientific correctness, comprehensibility, and actionability), suggesting that ChatGPT provided better responses across multiple distinct components relevant for providing high-quality answers.

While recent studies have compared the generative performance of different chatbots (Havers et al., 2025), our study is, to the best of our knowledge, the first to directly evaluate ChatGPT's capabilities against human peers using a blinded rating procedure for assessing informational support in a personal training context. In the field of nutrition, Kirk et al. (Kirk et al., 2023) found that responses from ChatGPT to commonly asked nutrition questions received higher scores than those of Dutch dietitians. Interestingly, both our findings and those of Kirk et al. show that the scores of the individual grading criteria were similar to the overall scores and that the overall differences were not simply driven by extreme scores in one of the grading components (Kirk et al., 2023). Furthermore, an overlapping finding in both studies was that on no occasion for either the overall scores or any individual grading criteria were scores for ChatGPT lower than those provided by human counterparts (Kirk et al., 2023). The questions in both studies belonged to a variety of subject areas within their respective domains, which suggests that ChatGPT's capabilities are not limited to specific fields but rather are wide-reaching. This indicates a clear advantage over human professionals, whose expertise is usually confined to a limited number of areas. However, admittedly, we cannot draw definitive conclusions about this hypothesis due to the limited number of questions and therefore subject areas in our study.

LLMs for fitness and health knowledge acquisition

Our findings support the use of ChatGPT for knowledge acquisition in the fitness and health domain. We propose that ChatGPT and other LLMs could support health professionals in promoting commitment to physical activity and exercise guidelines by providing accurate and comprehensible responses to inquiries in real time. Parallels can be seen in other branches of health, such as medicine and nutrition. In medicine, LLMs have been proposed to be used for patient queries, educating medical students or patients, and managing chronic diseases (Thirunavukarasu et al., 2023)(Kurniawan et al., 2024). In nutrition, they could be used to support the nutritional needs of those looking to improve their diet (Bergling et al., 2025).

In recent years, the internet has become a key source of information in the fitness and health space (Tan and Goonawardene, 2017). Given the number of users that ChatGPT has attracted, combined with its ability to synthesize information from different sources and thus eliminate the need to look through multiple web pages, it is reasonable to assume that ChatGPT will be used for knowledge

acquisition by those looking to deepen their knowledge on exercise-related topics. Hence, validating the adequacy of ChatGPT in this role is an important contribution of our work, and our findings support that ChatGPT could be used to answer exercise-related questions that might otherwise have been sought online or asked to a fitness professional.

AI-driven chatbots may support fitness professionals and trainees

Studies conducted so far in the field of exercise and health indicate that chatbots can improve the efficiency of resource allocation by offloading human operators from duties that can be automated (Fadhil and Gabrielli, 2017) and may enhance physical activity levels (Oh et al., 2021). This could provide opportunities for both PTs and trainees. For example, unlike PTs, chatbots are available 24/7, can generate answers instantly, and may be free of charge. Additionally, by responding to relatively basic questions, chatbots can free up PTs' time to focus on more interesting activities, particularly those for which chatbots are not well-suited. Initially, chatbots may primarily address simple questions, as in our study, but they could evolve to fulfill additional roles, such as functioning as virtual PTs and offering personalized approaches (Kirk et al., 2021). In these expanded roles, human PTs would still contribute essential elements, including direct interaction and emotional support.

Strengths

We highlight some strengths of our work. First, all participants—including PTs submitting questions and answers, grading PTs, and scientific experts—were blinded to group allocation, reducing bias. Second, the questions used were directly representative of those commonly asked by trainees seeking support from a PT, enhancing real-world relevance. Third, evaluating answers across three components (scientific correctness, comprehensibility, and actionability) provided a multidimensional assessment of answer quality. The grading structure – one-third scientific experts and two-thirds PTs – reflected these components appropriately, and the resulting grades demonstrated good inter-rater reliability.

Limitations

We note the following limitations. Firstly, although the sample size was smaller than initially planned (9 vs 17), the experimental design generated a total of 486 individual ratings, allowing stable comparative estimates. We therefore interpret the findings as exploratory and hypothesis-generating rather than definitive. Due to the limited sample size, we were unable to examine whether answer quality differed among PTs with varying levels of academic education. This is an important avenue for future research, particularly given the relatively low entry requirements for becoming a licensed PT. The inter-rater reliability (IRR) analysis among the scientific experts could only be conducted for those grading the five questions related to the "training for fat loss" topic, as for all other topics, each expert rated only one subset of answers. Given the rapid evolution of LLMs, it is reasonable to expect that newer versions of ChatGPT (e.g., GPT-5 and beyond) will demonstrate greater consistency and accuracy, particularly regarding factual and practical aspects of exercise prescription. Future research should therefore consider model versioning as a key methodological factor, systematically recording the model type, release date, and prompting protocol, enabling temporal benchmarking of AI performance.

Conclusion

Our study found that ChatGPT (version 3.5) outperformed human PTs in answering commonly asked exercise training questions. The overall quality of responses from ChatGPT was higher in 6/9 of the questions investigated, while scores from the answers of human PTs were higher in none of the questions. These findings also extended to each of the individual metrics used in the assessment of the quality of the answers, showing a general superiority of ChatGPT in providing actionable, comprehensible, and scientifically correct answers to common exercise training questions. Our results provide evidence that AI-driven LLMs such as ChatGPT may be used by knowledge-seeking trainees to answer common training questions, which could both enhance knowledge acquisition and thus encourage commitment to exercise guidelines as well as reduce the workload for PTs. Future work should look to validate our findings using a larger sample of questions and answers to better identify the strengths and weaknesses of LLMs such as ChatGPT across a broad range of exercise training topics.

Acknowledgements

The author(s) reported there is no funding associated with the work featured in this article. No potential conflict of interest was reported by the authors. Experiments comply with the current laws of the country in which they were performed. The data that support the findings of this study are available on request from the corresponding author.

References

Bergling, K., Antin, P., Hannun, Y. A., Rahmim, A., and Afshin, A. (2025) From bytes to bites: application of large language models to enhance nutritional recommendations. *Clinical Kidney Journal* **18(4)**, sfaf082. https://doi.org/10.1093/ckj/sfaf082

Dergaa, I., Abdulrashid, H., Abdelrahman, H., and Souissi, A. (2024) Using artificial intelligence for exercise prescription in personalised health promotion: A critical evaluation of OpenAI's GPT-4 model. *Biology of Sport* **42(2)**, 221-241. https://doi.org/10.5114/biolsport.2024.133661

Düking, P., Steinacker, J. M., Badtke, F., and Sperlich, B. (2024) ChatGPT generated training plans for runners are not rated optimal by coaching experts, but increase in quality with additional input information. *Journal of Sports Science and Medicine* **23(1)**, 56-72. https://doi.org/10.52082/jssm.2024.56

Fadhil, A. and Gabrielli, S. (2017) Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 261-265. https://doi.org/10.1145/3154862.3154914

Gabarron, E., Dorronzoro-Zubiete, E., Bradway, M., and Årsand, E. (2020) What do we know about the use of chatbots for public health? In *Digital Personalized Health and Medicine*, 796-800. https://doi.org/10.3233/SHTI200270

Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T. Y., White, L. F., Yinusa-Nyahkoon, L., Jack, H. E., Bickmore, T. W., and Lestoquoy, A. S. (2017) Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling* **100(9)**, 1720-1729. https://doi.org/10.1016/j.pec.2017.04.015

Han, R., Burke, L. E., and Sereika, S. M. (2023) Feasibility and acceptability of chatbots for nutrition and physical activity health promotion among adolescents: Systematic scoping review with adolescent consultation. *JMIR Human Factors* **10(1)**, e43227. https://doi.org/10.2196/43227

Havers, T., O'Connor, D., Dergaa, I., and Souissi, A. (2025) Reproducibility and quality of hypertrophy-related training plans generated by GPT-4 and Google Gemini as evaluated by coaching experts. *Biology of Sport* **42(2)**, 289-329. https://doi.org/10.5114/biolsport.2025.145911

Kirk, D., Catal, C., and Tekinerdogan, B. (2021) Precision nutrition: A systematic literature review. *Computers in Biology and Medicine* **133**, 104365. https://doi.org/10.1016/j.compbiomed.2021.104365

Kirk, D., van Eijnatten, E., and Camps, G. (2023) Comparison of answers between ChatGPT and human dieticians to common nutrition questions. *Journal of Nutrition and Metabolism* **2023(1)**, 5548684. https://doi.org/10.1155/2023/5548684

Koo, T. K. and Li, M. Y. (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* **15(2)**, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kurniawan, M. H., Al Mamun, M., and Sutanto, F. (2024) A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness. *Annals of Medicine* **56(1)**, 2302980. https://doi.org/10.1080/07853890.2024.2302980

Lupton, D. (2020) 'Better understanding about what's going on': young Australians' use of digital technologies for health and fitness. *Sport, Education and Society* **25(1)**, 1-13. https://doi.org/10.1080/13573322.2018.1555661

Malik, A. (2023) OpenAI's ChatGPT now has 100 million weekly active users. *TechCrunch*. https://techcrunch.com/2023/11/06/openais-chatgpt-now-has-100-million-weekly-active-users/

Moore, O. (2023) How are consumers using generative AI? *Andreessen Horowitz*. https://a16z.com/how-are-consumers-using-generative-ai/ (Accessed: September 10, 2024)

Morreel, S., Verhoeven, V., and Mathysen, D. (2023) Microsoft Bing outperforms five other generative artificial intelligence chatbots in the Antwerp University multiple choice medical license exam. *medRxiv*, 2023.08.18.23294263. https://doi.org/10.1101/2023.08.18.23294263

Oh, Y. J., Zhang, J., Fang, M. L., and Xu, Y. (2021) A systematic review of artificial intelligence chatbots for promoting physical activity, healthy diet, and weight loss. *The International Journal of Behavioral Nutrition and Physical Activity* **18(1)**, 160. https://doi.org/10.1186/s12966-021-01224-6

Piao, M., Ryu, H., Lee, H., Kim, J., and Lee, J. (2020) Use of the healthy lifestyle coaching chatbot app to promote stair-climbing habits among office workers: Exploratory randomized controlled trial. *JMIR mHealth and uHealth* **8(5)**, e15085. https://doi.org/10.2196/15085

Ray, P. P. (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* **3**, 121-154. https://doi.org/10.1016/j.iotcps.2023.04.003

Tan, S. S.-L. and Goonawardene, N. (2017) Internet health information seeking and the patient-physician relationship: A systematic review. *Journal of Medical Internet Research* **19(1)**, e5729. https://doi.org/10.2196/jmir.5729

Thirunavukarasu, A. J., Elmas, B., Ramachandran, N., and Panch, T. (2023) Large language models in medicine. *Nature Medicine* **29(8)**, 1930-1940. https://doi.org/10.1038/s41591-023-02448-8

Washif, J. A., Ramlan, A. A., and Wong, M. Y. (2024) Artificial intelligence in sport: Exploring the potential of using ChatGPT in resistance training prescription. *Biology of Sport* **41(2)**, 209-220. https://doi.org/10.5114/biolsport.2024.132987

Zaleski, A. L., Thomas, R. J., Allison, K. M., and Arena, R. (2024) Comprehensiveness, accuracy, and readability of exercise recommendations provided by an AI-based chatbot: Mixed methods study. *JMIR Medical Education* **10**, e51308. https://doi.org/10.2196/51308

Zhu, W., Liu, M., Wang, H., and Gao, Z. (2024) Who could and should give exercise prescription: Physicians, exercise and health scientists, fitness trainers, or ChatGPT? *Journal of Sport and Health Science* **13(3)**, 368-372. https://doi.org/10.1016/j.jshs.2024.01.001

Key points

- ChatGPT outperformed personal trainers in answering six of nine exercise training-related questions.
- Scores were consistently higher for ChatGPT across all metrics measured, including scientific correctness, actionability and comprehensibility.
- Large language models may be useful tools to support personal trainers and trainees by providing comprehensible and scientifically correct answers to frequently asked exercise training questions.

AUTHOR BIOGRAPHY



Brecht D'HOE
Employment
Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium.
Degree
MSc
Research interests
Nutrigenetics, applied resistance training
E-mail: brecht.dhoe@ugent.be



Daniel KIRK
Employment
Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas Hospital, Westminster Bridge Road, London SE1 7EH, UK.
Degree
MSc
Research interests
Microbiome, Nutrition, Personalized Nutrition, Machine Learning
E-mail: daniel.1.kirk@kcl.ac.uk



Jan BOONE
Employment
Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium
Degree
PhD, Professor
Research interests
Physical conditioning, Sports Training, Sports Physiology
E-mail: jan.boone@ugent.be



Alessandro COLOSIO
Employment
Inter-University Laboratory of Human Movement Biology, Université Saint-Etienne, Saint-Etienne, France
Degree
PhD, Professor
Research interests
Exercise Physiology, Exercise Prescription, Exercise Therapy
E-mail: alessandro.colosio@univ-st-etienne.fr

✉ **Brecht D'hoe**
Department of Movement and Sports Sciences, Ghent University, Watersportlaan 2, Ghent, 9000, Belgium

SUPPLEMANTRY TABLES

Table S1. Additional personal training-related qualifications that the nine contributing personal trainers included in the study obtained.

	Additional personal training related qualifications
PT 1	BodyBuilding Trainer (Olympia) - Olympics Weightlifting Trainer Level 1
PT 2	Teacher Training Certification Program In the Pilates method
PT 3	None
PT 4	Degree in Sports, Crossfit level 1, Bodybuilding course, personalized training course, TRX course, Strength Development (Les mills), specialization course for pregnant women
PT 5	FMS, Strongfirst, Menno Henselmans, Backfitpro, EBFA, Nick Tumminello
PT 6	Certificate in Fitness Instructing, and further qualifications in Functional Movement Systems, Optimum Performance, Bootcamps, Circuits, Studio Cycling, Water Fitness, Behaviour Change Science, Mental Toughness, Principles of Nutrition, and Plant-based Diets.
PT 7	Various exercise and science related certifications and CPD courses
PT 8	None
PT 9	Basic medical knowledge, Orthomolecular therapist

Table S2. Country of employment and current function of scientific expert graders.

Grader nr.	Grader nr.	Country of employment	Current function
Question topic: Training frequency	1	Canada	Professor in kinesiology
	2	Australia	Professor in sports physiology & strength and conditioning
	3	Norway	Professor in exercise and strength training
	4	Spain	Professor in exercise intervention for health
	5	Australia	Senior lecturer in exercise physiology and performance
	6	Spain	Professor in physical education and sports
	7	Australia	Senior lecturer in exercise physiology and nutrition research
	8	Australia	Lecturer in neurophysiology and resistance training
	9	Finland	Academy research fellow in neuromuscularity and resistance training
Question topic: Training for fat loss	1	USA	Professor in endocrinology and metabolism
	2	Greece	Professor in exercise biochemistry
	3	Spain	Substitute teacher in sports medicine
	4	USA	Professor in health sciences
	5	UK	Professor in psychobiology
	6	USA	Professor in kinesiology and nutrition sciences
	7	USA	Professor in nutrition and exercise physiology
	8	USA	Professor in exercise science
	9	New Zealand	Senior research fellow in sport & exercise science
Question topic: Training motivation	1	Portugal	Professor in endocrinology and metabolism
	2	UK	Professor in motivational science and health behavior change
	3	Canada	Professor in sports psychology
	4	USA	Professor in pedagogical sciences
	5	New Zealand	Professor in sports psychology
	6	Belgium	Postdoctoral researcher in Sports psychology
	7	UK	Senior Lecturer in coaching science
	8	UK	Reader in sports and exercise psychology
	9	UK	Professor in sports coaching
Question topic: Training around pain	1	Norway	Professor in sports trauma
	2	UK	Professor in physiotherapy and rehabilitation sciences
	3	UK	Senior research fellow in spinal physiotherapy
	4	Denmark	Professor in muscle and joint health
	5	Ireland	Professor in physiotherapy
	6	Belgium	Professor in revalidation sciences and physiotherapy
	7	Ireland	Professor in physiotherapy
	8	Australia	Post doctoral research fellow in sports and exercise physiotherapy
	9	Germany	Professor in training intervention research
Question topic: Timing of the day for training	1	USA	Professor in medicine, endocrinology; diabetes and metabolism
	2	the Netherlands	Professor in human biology
	3	UK	Lecturer in exercise physiology and diabetes
	4	USA	Professor in kinesiology and nutrition sciences
	5	USA	Professor in biological timing
	6	Australia	Senior lecturer in exercise and sports sciences
	7	Canada	Professor in pediatrics and obesity
	8	the Netherlands	Professor in metabolism and diabetes
	9	USA	Professor in human physiology

Table S3. Nationality and obtained qualifications of grading PTs.

Grader nr.	Nationality	Obtained qualifications
1	Ukrainian	EREPS level 4 personal trainer, EREPS master trainer
2	Italian	EREPS level 4 personal trainer, EREPS master trainer
3	Canadian	Canfitpro certified personal trainer
4	Belgian	Bachelor's degree in physical education and recreation, specializing in personal training
5	Belgian	EREPS level 4 personal trainer
6	Belgian	Bachelor's degree in nutrition and dietetics, Calisthenics trainer A, Calisthenics trainer B, YPSI advanced programming
7	Belgian	Bachelor's degree in physical education and movement sciences
8	Belgian	Bachelor's degree in physical education and recreation, specializing in fitness and health
9	Belgian	Master's degree in rehabilitation sciences, bachelor's degree in nutrition and dietetics, Trainer B badminton
10	Belgian	NASM personal trainer's degree, NSCA personal trainer's degree, EXOS degree
11	Belgian	Bachelor's degree in physical education and recreation, Postgraduate sportmanagment, Postgraduate advanced personal training
12	Belgian	NASM personal trainer's degree
13	Portuguese	EREPS level 4 personal trainer
14	Belgian	EREPS level 4 personal trainer
15	Belgian	EREPS level 4 personal trainer
16	Belgian	EREPS level 4 personal trainer, specialization in EMS-training
17	Belgian	EREPS level 4 personal trainer, NASM personal trainer's degree
18	Belgian	EREPS level 4 personal trainer

Table S4. Summary statistics for average overall scores.

Question Number		Mean	Median	Interquartile Range	Minimum	Maximum
1	ChatGPT	7.57	7.67	1.83	4.00	10.00
1	PT	6.69	6.67	2.50	3.67	10.00
2	ChatGPT	7.96	8.33	1.83	4.33	10.00
2	PT	5.49	5.33	2.33	2.00	10.00
3	ChatGPT	7.98	8.33	1.67	4.33	10.00
3	PT	7.01	7.33	2.17	3.33	10.00
4	ChatGPT	7.40	7.33	1.33	4.00	10.00
4	PT	7.44	7.67	1.83	3.67	10.00
5	ChatGPT	8.11	8.17	0.13	6.67	8.42
5	PT	7.78	8.00	1.50	5.00	9.33
6	ChatGPT	7.99	7.67	1.50	5.00	10.00
6	PT	7.07	7.33	2.00	3.00	10.00
7	ChatGPT	8.15	8.67	1.33	5.00	10.00
7	PT	6.36	6.67	2.83	2.33	10.00
8	ChatGPT	7.73	7.67	2.00	4.33	10.00
8	PT	6.65	7.00	2.83	2.67	9.33
9	ChatGPT	7.54	7.67	1.83	3.33	10.00
9	PT	6.98	7.00	2.17	3.33	10.00

Table S5. Answers given to each of the questions by ChatGPT and personal trainers.

Question

How often do I need to train for the training to be effective?

ChatGPT

The effectiveness of training depends on several factors, including the type of training, your goals, and your current fitness level. For general fitness and health, most guidelines recommend a mix of cardiovascular, strength, and flexibility training throughout the week.

- Cardiovascular Training: Aim for at least 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity aerobic activity per week, spread across several days. This could include activities like brisk walking, running, cycling, or swimming.
- Strength Training: Include two to three strength training sessions per week, focusing on major muscle groups. This helps improve muscle tone, strength, and endurance.
- Flexibility and Mobility: It's also beneficial to include stretching and mobility exercises in your routine to improve flexibility and reduce the risk of injury. Aim for flexibility training two to three times a week.

You can tailor the frequency and intensity of your training based on your personal goals and fitness level. Consistency is key; it's better to train consistently at a moderate level than to push too hard and risk burnout or injury. Always listen to your body and adjust your training accordingly.

Personal trainer 6

At least 3 times a week to progress. If you train twice a week or less you will not progress but maintain your fitness level, which ever aspect of it your training is targeting. If you're a beginner, aim for 3 sessions per week, allowing at least one day of rest between

sessions to allow muscles to recover. If you're intermediate or advanced, you can increase frequency to 4 to 6 times per week, targeting different muscle groups, and cardio systems, on different days. Make sure to get a rest day or a low intensity day (doing yoga or the likes) every week to avoid strain and injuries.

Question
How do I lose abdominal fat?

ChatGPT
Losing abdominal fat involves a combination of dietary changes, regular exercise, and lifestyle adjustments. Here are some strategies to help you lose abdominal fat:

1. Healthy Diet: Focus on a balanced diet rich in fruits, vegetables, whole grains, lean proteins, and healthy fats. Limit your intake of refined sugars, processed foods, and high-calorie snacks.
2. Caloric Deficit: To lose fat, you need to consume fewer calories than you burn. Monitor your caloric intake and make adjustments as needed to create a caloric deficit.
3. Regular Exercise: Incorporate both cardiovascular exercise (such as running, cycling, or swimming) and strength training (such as weight lifting) into your routine. Cardio helps burn calories, while strength training helps build muscle and increase metabolism.
4. Core Exercises: Include exercises that target your core muscles, such as planks, crunches, and other ab workouts. These can help strengthen your abdominal muscles.
5. Stay Hydrated: Drinking enough water can help you stay hydrated and feel full, which may help you avoid overeating.
6. Manage Stress: High stress levels can lead to weight gain, particularly around the abdomen. Practice stress-reducing techniques such as meditation, yoga, or deep breathing exercises.
7. Get Enough Sleep: Aim for 7-9 hours of quality sleep per night. Poor sleep can disrupt hormones that regulate appetite and metabolism, making it harder to lose fat.
8. Consistency: Be consistent with your diet and exercise routine. It takes time to see results, so stick with your plan and be patient.

Remember that spot reduction (targeting fat loss in a specific area) is not possible. However, by following these strategies, you can work toward overall fat loss, including in the abdominal area.

Personal trainer 1
First I ask if the client has any hormonal pathology; If it is positive, this pathology must first be normalized with an endocrine doctor. Secondly, I review your eating style to make adjustments to your diet to help achieve this goal. Third, if possible, I prescribe intermittent fasting with a healthy diet. Fourth, I prescribe the training program according to the client's available time, the volume of work should be 70% cardio 30% strength and core circuits in each session.

Question
How can I maintain my motivation to work out?

ChatGPT
Maintaining motivation to work out can sometimes be a challenge, but there are several strategies you can use to stay consistent and enthusiastic about your fitness routine. One effective approach is to set clear, achievable goals for yourself. Start with small, manageable goals that are easy to achieve and gradually build up to more challenging ones. This will help you experience a sense of accomplishment along the way.
Another key factor is finding a workout routine that you enjoy. Whether it's running, yoga, weightlifting, or dancing, choosing activities that you genuinely like can make a big difference in your motivation. Additionally, mixing up your routine can keep things interesting and prevent boredom.
Establishing a consistent schedule for your workouts can also help you stay motivated. Try to set aside a specific time each day or week for exercise, and treat it as a regular appointment with yourself. This can help create a habit and make it easier to stick with your routine.
Tracking your progress is another great way to stay motivated. Keep a workout journal or use a fitness app to monitor your workouts, achievements, and any improvements over time. Seeing your progress can boost your confidence and inspire you to keep going.
Finally, find ways to reward yourself for sticking with your routine. Treat yourself to something you enjoy, like a relaxing spa day, a favorite snack, or a new workout outfit. Rewards can provide positive reinforcement and give you something to look forward to as you continue your fitness journey.

Personal trainer 2
Firstly, by creating a long term goal. An example would be, the client wants to gain strength and mobility within 1 year. I will create a workout plan that includes mobility exercises for a warm-up, strength with sets, reps and weight, and cool down which will incorporate stretching. Secondly, breaking that goal down into smaller short term goals. An example would be, every three months, the client and I will meet to discuss their outcomes, challenges and progression? I will modify the workout plan according to the next phase and set out actionable steps to meet the next bite size goal and meet again. I will hold them accountable by asking them to send me a challenge that they would like to create for themselves as a form of evidence. This could be a photograph of how many sets, reps and weight that they would like to achieve or how many steps they have recorded during the day or even weekly. It is important to always have some form of contact with the client, to remind them. Either, WhatsApp, or message or social media or even seeing them in the gym environment and acknowledging them.

Question
How do I train to lose weight?

ChatGPT
Training to lose weight involves a combination of exercise, healthy eating habits, and consistency. Here's how you can approach your weight loss training:

1. Cardiovascular Exercise: Cardio exercises such as running, cycling, swimming, or using the elliptical machine help burn calories and improve heart health. Aim for at least 150 minutes of moderate-intensity cardio each week.
2. Strength Training: Incorporating strength training exercises, such as lifting weights or using resistance bands, helps build muscle and increase your metabolism. This can help you burn more calories even when you're resting.
3. High-Intensity Interval Training (HIIT): HIIT involves alternating between short bursts of intense activity and periods of rest or lower-intensity activity. This type of training can be very effective for burning calories and improving fitness.
4. Flexibility and Mobility Exercises: Stretching and yoga can help improve flexibility, reduce injury risk, and keep your body balanced.
5. Healthy Eating Habits: Focus on a balanced diet with plenty of vegetables, fruits, whole grains, lean proteins, and healthy fats. Pay attention to portion sizes and aim to create a calorie deficit by consuming fewer calories than you burn.
6. Stay Hydrated: Drinking plenty of water helps regulate your body temperature, keeps you hydrated, and may help control your appetite.
7. Consistency: Consistency is key when it comes to losing weight. Create a routine that you can stick to over time.
8. Rest and Recovery: Allow your body time to recover by getting enough sleep and giving yourself rest days from intense exercise.

Remember to set realistic goals and track your progress. It's important to listen to your body and make adjustments as needed.

Personal trainer 3
First and foremost, you need to be in a caloric deficit. That means expending more energy than you consume everyday. We can estimate how many calories we need to maintain our current weight through the use of a total daily energy expenditure (TDEE) calculator. TDEE is the amount of calories we burn from our resting metabolism plus our activities daily. Then, the goal is to consume calories that is less than this number. I would recommend a deficit of around 300-500 kcal less than the TDEE. This is to ensure we still get enough nutrition to maintain wellbeing and not be starving constantly. And of course, leading an active lifestyle will bring up the TDEE which makes it easier to stay in a caloric deficit. Keep this up consistently and you will definitely lose weight. Now as to how we can incorporate exercise, what we really want is for all the weight loss to be from fat and not muscle. This is because muscle is more metabolically active than fat and greatly contributes to our resting metabolism rate. Having more muscle also results in more strength which allows us to be more intense in our workouts. Thus, we should primarily do resistance training for strength in a progressively more challenging manner each session in order to maintain or grow muscle while the fat burns away. 2 to 3 times per week would suffice. As for cardio training, it will help in burning calories for sure but the main reason to do them should be to maintain heart health. You may do cardio as often as you like at an intensity that does not compromise your performance during your strength training.

Question
Is cardio better to lose weight than strength workouts?

ChatGPT
Both cardio and strength workouts can play a role in weight loss, but they work in different ways and have different benefits. Cardio exercises, such as running, cycling, and swimming, burn a significant amount of calories during the workout and can help you lose weight. These activities increase your heart rate and help improve cardiovascular health.
On the other hand, strength training exercises, like lifting weights or using resistance bands, help you build muscle mass. While these workouts may not burn as many calories during the exercise as cardio, building muscle can increase your resting metabolic rate. This means you may burn more calories even when you are at rest, which can contribute to weight loss over time.
A combination of both cardio and strength training can be the most effective approach for weight loss and overall fitness. Cardio workouts can help you burn calories and improve heart health, while strength training can help you build muscle and increase your metabolism. Incorporating both types of exercise into your routine can help you achieve a balanced fitness regimen and optimize your weight loss efforts.

Personal trainer 4
Cardio exercises, like running or cycling, are excellent for burning calories and promoting weight loss in the short term because they get your heart rate up and help you burn more calories. However, it's important to note that strength training, like lifting weights, is also crucial for weight loss. Research conducted by Paulo Gentil, among others in the field of exercise science, has shown that combining strength training with cardio can lead to even greater reductions in body fat compared to just doing cardio alone. This means that incorporating a variety of exercises into your routine, including both cardio and strength workouts, is the best approach for achieving your weight loss goals and improving overall health.

Question
How can I train around pain?

ChatGPT
Training around pain involves finding ways to adjust your workouts to accommodate your body's needs while avoiding further injury. Start by identifying the source and type of pain you are experiencing. If the pain is acute or severe, it's crucial to consult a healthcare professional to rule out serious injuries or conditions.
When modifying your exercise routine, focus on these strategies:

1. Modify Exercises: Substitute exercises that cause pain with alternatives that target the same muscle groups without discomfort. For example, if running hurts your knees, consider swimming or cycling as lower-impact alternatives.
2. Adjust Intensity: Lower the intensity, duration, or frequency of your workouts. Gradually increase as you gain strength and confidence, being mindful of your pain levels.
3. Warm Up Properly: Take extra time to warm up before exercising, which can help prepare your body and reduce the risk of pain or injury.
4. Use Support: Utilize supportive gear such as braces, wraps, or insoles to help stabilize areas prone to pain.
5. Listen to Your Body: Pay attention to your body's signals and stop any activity that exacerbates your pain.
6. Incorporate Rest Days: Rest is essential for recovery, especially if you're dealing with pain. Be sure to include rest days in your training schedule.
7. Seek Guidance: Consider working with a physical therapist who can tailor their guidance to your specific pain.

Adjusting your training regimen with these strategies can help you stay active while avoiding additional pain or injury.

Personal trainer 5
When facing pain, seeking guidance from a certified physiotherapist is paramount to understanding the underlying issues and receiving appropriate treatment recommendations. Following their professional assessment, I can provide tailored guidance on effective strength training techniques that align with your specific needs and rehabilitation goals. Central to this approach is prioritizing proper form and technique to avoid exacerbating discomfort and promoting safe, gradual progress.
Moreover, it's essential to underscore the significance of patience and consistency in adhering to the prescribed protocol. Recovery from injury often requires diligence over time, and rushing the process can lead to setbacks. By embracing a mindset of steady, incremental improvement and respecting your body's signals, you can optimize the rehabilitation journey.
Ultimately, the goal is not just to alleviate pain but to rebuild strength, resilience, and overall physical well-being. With a collaborative effort between professional guidance and dedicated personal commitment, you can navigate through injury recovery effectively and emerge stronger on the other side.

Question
Working out at what time of the day leads to most effective results?

ChatGPT
The best time of day to work out can vary depending on individual preferences, lifestyle, and fitness goals. While there isn't a definitive answer that applies to everyone, there are a few factors to consider that can help you determine the best time for you:

1. Energy Levels: Some people feel more energetic and motivated to work out in the morning, while others prefer the afternoon or evening when they've had time to eat and hydrate throughout the day.
2. Consistency: The best time to work out is the time you can consistently stick to. If you find it easier to schedule exercise in the morning before work or in the evening after your responsibilities, that might be the best choice for you.
3. Performance: Some research suggests that exercise performance may peak in the late afternoon or early evening when body temperature is higher. This can lead to better strength and endurance during your workout.
4. Sleep Patterns: Morning workouts can help set a positive tone for the day and may improve sleep quality at night. Evening workouts can help release stress accumulated during the day.
5. Goals: Your specific fitness goals may also influence your preferred workout time. For example, if you're training for a morning race, practicing at a similar time may be beneficial.

Ultimately, the most effective time to work out is the time that works best for your schedule and allows you to maintain a consistent routine. Try experimenting with different times of the day to see which suits you best.

Personal trainer 7
Well, it all depends on your routine, if you are a very busy professional, and happen to have a regular time availability at the start of the day, then it would be best to promote healthier habits leading to self-discipline if you are considering health and fitness by maintenance, and with the passage of time it may also boost exercise performance. On the other hand, if the focus is on strength training or conditioning, then an evening workout is preferred. In any of the cases, it is most important to consider the role of recovery via sleep versus intensity or strain of activities performed. For most individuals who are diurnal, may find it easy to recover via night sleep.
In short, if time allows, it is best to exercise at the time when you feel the most active in a 24-hour period.

Question
Which exercises are best for losing weight?

ChatGPT
Exercise can play a significant role in weight loss when combined with a healthy diet and consistent routine. Some of the best exercises for losing weight include a mix of cardiovascular activities, strength training, and high-intensity interval training (HIIT):

1. Cardiovascular Exercises: Running, cycling, swimming, and walking are great for burning calories and improving heart health. These activities can be done outdoors or on gym equipment like treadmills and stationary bikes.

2. Strength Training: Lifting weights or using resistance bands helps build muscle mass, which can increase your resting metabolic rate and help you burn more calories even at rest. Focus on compound exercises like squats, deadlifts, and bench presses for the best results.
3. High-Intensity Interval Training (HIIT): This form of exercise alternates between short bursts of intense activity and periods of rest or lower-intensity activity. HIIT can boost your metabolism and lead to significant calorie burn in a short amount of time.
4. Circuit Training: Combining strength training and cardiovascular exercises in a circuit format can be highly effective for weight loss. This involves performing a series of exercises back-to-back with little rest in between.
5. Aerobic Dance Classes: Classes like Zumba or other dance-based workouts can be a fun and effective way to burn calories.
6. Bodyweight Exercises: Exercises like push-ups, squats, and lunges can be done anywhere and can help tone your body while burning calories.

The key to successful weight loss through exercise is consistency and finding a routine that you enjoy and can maintain over time. Make sure to pair your exercise routine with a healthy, balanced diet for optimal results.

Personal trainer 8

To lose weight, you want to achieve a significant calorie burn. It's best to choose major lifts such as the deadlift, barbell back squat, and bench press. These are compound exercises that train multiple muscle groups simultaneously and will require more recovery after a workout than if you only worked out smaller muscle groups. Additionally, you can train other muscle groups as well, depending on your priorities. I also recommend supplementary cardio and, of course, paying attention to your diet.

Question

How do I get a tighter stomach?

ChatGPT

Achieving a tighter stomach usually involves a combination of proper diet, regular exercise, and overall lifestyle changes. Here's what you can do to work towards your goal:

1. Healthy Diet: Focus on a balanced diet with an appropriate amount of calories for your goals. Include plenty of vegetables, fruits, lean proteins, whole grains, and healthy fats. Reducing your intake of sugary drinks and processed foods can also help.
2. Strength Training: Incorporate exercises that target your core muscles, such as planks, sit-ups, leg raises, and Russian twists. These exercises help strengthen your abdominal muscles.
3. Cardiovascular Exercise: Activities such as running, cycling, swimming, or even brisk walking can help you burn calories and reduce body fat, including the fat around your stomach.
4. Consistency: Consistency is key. It can take time to see changes in your stomach, so stick to your exercise and diet plan.
5. Hydration: Drink plenty of water throughout the day. Staying hydrated can help control hunger and aid in digestion.
6. Sleep: Getting enough quality sleep is important for your overall health and can help regulate hormones that affect appetite and fat storage.
7. Stress Management: High levels of stress can lead to weight gain around the stomach area due to increased levels of the hormone cortisol. Managing stress through techniques such as meditation, yoga, or deep breathing can help.

Everyone's body is different, so results may vary. Always listen to your body and adjust your routine as needed.

Personal trainer 9

You don't get a tighter stomach simply by doing abdominal exercises; you achieve it by training and nourishing the whole body. It's important to stimulate large muscle groups such as the glutes, hamstrings, quadriceps, latissimus, and pectoralis in this process. This is because activating and developing these muscles through training increases energy expenditure and fat burning, making it easier to target the fat tissue around the stomach. Additionally, eating a healthy diet is also crucial in this process, including an adequate intake of proteins, healthy fats, and carbohydrates (as well as vitamins and minerals) in the right proportions.

Table S6. All grades given by ChatGPT and personal trainers to all nine questions

Scientific correctness Q1 ChatGPT	Comprehensibility Q1 ChatGPT	Actionability Q1 ChatGPT	Overall score Q1 ChatGPT
Scientific experts			
8	8	8	8
9	8	9	9
8	9	7	8
8	9	9	9
8	8	7	8
7	10	10	9
8	8	6	7
7	9	9	8
8	9	8	8
PTs			
9	8	7	8
8	10	9	9
9	8	5	7
9	5	2	5
9	6	5	7
8	8	4	7
8	4	7	6
9	8	6	8
9	5	4	6
10	10	10	10
6	3	3	4
9	9	9	9
9	9	8	9
8	7	9	8
9	7	7	8
10	7	3	7
8	8	7	8
5	7	5	6
Scientific correctness Q2 ChatGPT	**Comprehensibility Q2 ChatGPT**	**Actionability Q2 ChatGPT**	**Overall score Q2 ChatGPT**
Scientific experts			
7	7	7	7
10	10	10	10
8	10	10	9
9	9	9	9
5	6	2	4
8	9	6	8
9	9	8	9
6	9	6	7
8	8	10	9
PTs			
10	9	8	9
10	10	10	10
7	7	5	6
8	5	4	6
9	9	7	8
4	9	6	6
7	8	2	6
9	10	7	9
8	8	7	8
9	10	10	10
8	8	8	8
9	9	9	9
7	9	7	8
7	8	8	8
9	9	8	9
10	10	5	8
8	8	6	7
8	10	10	9

Scientific correctness Q3 ChatGPT	Comprehensibility Q3 ChatGPT	Actionability Q3 ChatGPT	Overall score Q3 ChatGPT
Scientific experts			
8	8	8	8
6	10	9	8
8	8	9	8
9	9	9	9
8	8	8	8
7	10	10	9
6	7	7	7
8	9	9	9
7	7	7	7
PTs			
8	10	9	9
10	10	8	9
7	8	7	7
1	7	7	5
8	8	9	8
8	6	6	7
6	7	8	7
7	10	8	8
6	8	8	7
9	9	9	9
7	10	9	9
9	9	9	9
8	8	7	8
7	9	8	8
9	9	9	9
8	10	8	9
1	5	7	4
10	10	10	10
Scientific correctness Q4 ChatGPT	**Comprehensibility Q4 ChatGPT**	**Actionability Q4 ChatGPT**	**Overall score Q4 ChatGPT**
Scientific experts			
6	7	6	6
10	10	10	10
8	9	6	8
8	8	8	8
6	6	3	5
9	8	4	7
8	6	7	7
6	8	6	7
7	10	10	9
PTs			
8	7	6	7
9	9	8	9
8	8	3	6
9	3	0	4
9	8	6	8
7	8	6	7
7	7	7	7
10	8	7	8
7	8	7	7
9	9	9	9
7	7	7	7
9	9	9	9
7	9	6	7
9	7	7	8
9	8	6	8
6	8	5	6
6	6	7	6
8	10	10	9

Scientific correctness Q5 ChatGPT	Comprehensibility Q5 ChatGPT	Actionability Q5 ChatGPT	Overall score Q5 ChatGPT
Scientific experts			
4	5	3	4
8	10	10	9
6	8	8	7
9	9	8	9
8	7	6	7
9	9	5	8
9	9	9	9
9	8	7	8
8	10	6	8
PTs			
10	8	8	9
10	10	10	10
7	8	7	7
2	8	5	5
10	8	8	9
7	8	7	7
8	8	6	7
8	10	9	9
9	8	8	8
10	10	9	10
6	8	8	7
9	9	9	9
8	9	8	8
8	8	8	8
10	10	7	9
10	10	8	9
8	8	8	8
5	5	10	7
Scientific correctness Q6 ChatGPT	**Comprehensibility Q6 ChatGPT**	**Actionability Q6 ChatGPT**	**Overall score Q6 ChatGPT**
Scientific experts			
9	9	10	9
5	6	7	6
6	8	8	7
7	8	8	8
5	7	8	7
8	9	9	9
7	8	8	8
9	9	10	9
8	8	6	7
PTs			
10	8	7	8
10	10	10	10
8	8	7	8
6	7	2	5
9	9	8	9
8	8	7	8
8	7	8	8
8	8	7	8
8	8	7	8
10	10	10	10
6	7	8	7
9	9	9	9
5	9	6	7
8	9	7	8
9	10	10	10
8	10	5	8
7	7	8	7
10	10	10	10

Scientific correctness Q7 ChatGPT	Comprehensibility Q7 ChatGPT	Actionability Q7 ChatGPT	Overall score Q7 ChatGPT
Scientific experts			
7	7	6	7
7	8	9	8
10	10	9	10
8	10	10	9
9	8	9	9
7	9	8	8
9	9	8	9
7	5	3	5
8	9	9	9
PTs			
8	7	7	7
10	10	10	10
9	9	8	9
4	6	5	5
9	9	8	9
7	8	6	7
7	8	8	8
9	10	10	10
7	9	8	8
10	10	10	10
7	9	9	8
9	9	9	9
5	9	7	7
7	6	7	7
9	9	8	9
9	9	8	9
8	8	8	8
7	10	10	9
Scientific correctness Q8 ChatGPT	**Comprehensibility Q8 ChatGPT**	**Actionability Q8 ChatGPT**	**Overall score Q8 ChatGPT**
Scientific experts			
5	5	4	5
10	10	10	10
9	8	10	9
9	9	9	9
8	7	4	6
8	8	6	7
8	7	7	7
7	8	6	7
8	10	8	9
PTs			
7	9	10	9
9	8	10	9
8	8	7	8
4	6	3	4
7	8	6	7
7	9	6	7
8	8	8	8
9	10	7	9
7	8	7	7
10	10	10	10
7	7	7	7
9	9	9	9
5	9	5	6
4	5	6	5
9	9	9	9
10	10	8	9
6	8	6	7
7	10	10	9

Scientific correctness Q9 ChatGPT	Comprehensibility Q9 ChatGPT	Actionability Q9 ChatGPT	Overall score Q9 ChatGPT
Scientific experts			
3	4	3	3
6	10	10	9
7	9	8	8
9	9	9	9
4	6	4	5
9	9	5	8
8	8	6	7
6	7	7	7
8	10	10	9
PTs			
7	8	9	8
10	10	10	10
8	8	6	7
6	5	4	5
9	9	6	8
8	8	6	7
8	7	7	7
10	7	7	8
6	7	7	7
9	9	9	9
7	7	7	7
9	9	9	9
5	9	6	7
6	7	7	7
9	8	7	8
8	10	7	8
6	7	7	7
10	10	10	10
Scientific correctness Q1 PT	**Comprehensibility Q1 PT**	**Actionability Q1 PT**	**Overall score Q1 PT**
Scientific experts			
3	5	5	4
6	6	7	6
2	4	6	4
8	10	8	9
4	7	7	6
8	10	10	9
4	4	8	5
7	7	7	7
6	8	8	7
PTs			
8	8	8	8
7	10	8	8
1	7	3	4
3	8	7	6
7	9	8	8
1	8	4	4
3	7	6	5
9	10	7	9
9	7	4	7
10	10	10	10
6	9	9	8
7	7	7	7
2	9	4	5
6	8	8	7
7	9	9	8
6	8	5	6
5	6	7	6
2	10	5	6

Scientific correctness Q2 PT	Comprehensibility Q2 PT	Actionability Q2 PT	Overall score Q2 PT
Scientific experts			
3	3	3	3
5	8	8	7
5	7	4	5
8	5	7	7
3	4	2	3
5	5	3	4
5	6	5	5
5	7	7	6
3	5	1	3
PTs			
5	6	5	5
10	10	10	10
6	6	4	5
3	3	3	3
9	5	6	7
3	7	4	5
4	2	3	3
7	6	6	6
2	8	6	5
8	9	9	9
4	1	1	2
7	7	7	7
4	9	5	6
6	8	8	7
4	8	7	6
7	10	5	7
5	4	4	4
5	7	5	6
Scientific correctness Q3 PT	**Comprehensibility Q3 PT**	**Actionability Q3 PT**	**Overall score Q3 PT**
Scientific experts			
5	5	4	5
6	8	8	7
8	6	9	8
8	9	9	9
8	8	8	8
9	8	10	9
5	5	4	5
4	6	4	5
6	5	8	6
PTs			
7	6	7	7
10	10	10	10
6	5	6	6
1	6	3	3
7	9	8	8
4	7	7	6
6	9	7	7
8	8	8	8
4	8	8	7
6	7	10	8
5	3	5	4
7	7	7	7
8	8	8	8
7	9	8	8
8	10	9	9
8	10	8	9
2	7	7	5
8	10	8	9

Scientific correctness Q4 PT	Comprehensibility Q4 PT	Actionability Q4 PT	Overall score Q4 PT
Scientific experts			
4	5	2	4
8	10	10	9
8	7	8	8
9	7	7	8
7	7	5	6
6	8	5	6
8	9	8	8
9	8	7	8
9	10	8	9
PTs			
10	4	6	7
10	10	10	10
8	7	5	7
8	5	5	6
7	9	9	8
7	7	7	7
9	7	6	7
9	8	9	9
6	6	6	6
10	9	10	10
8	3	7	6
7	7	7	7
8	9	8	8
9	8	8	8
9	8	6	8
5	5	5	5
7	7	8	7
8	8	10	9
Scientific correctness Q5 PT	**Comprehensibility Q5 PT**	**Actionability Q5 PT**	**Overall score Q5 PT**
Scientific experts			
5	6	4	5
7	7	7	7
9	9	9	9
9	9	8	9
8	7	6	7
9	8	5	7
8	8	7	8
7	8	7	7
10	10	6	9
PTs			
9	10	9	9
8	8	8	8
8	9	7	8
8	7	3	6
7	7	9	8
7	8	7	7
6	6	7	6
9	10	9	9
9	9	8	9
7	7	9	8
8	9	9	9
7	7	7	7
9	9	8	9
9	7	8	8
9	9	8	9
10	10	8	9
8	8	8	8
2	10	5	6

Scientific correctness Q6 PT	Comprehensibility Q6 PT	Actionability Q6 PT	Overall score Q6 PT
Scientific experts			
4	4	3	4
4	3	2	3
10	7	9	9
6	6	5	6
8	7	7	7
7	7	6	7
7	3	3	4
7	9	6	7
7	5	3	5
PTs			
10	6	7	8
10	9	8	9
9	8	8	8
2	7	4	4
9	7	6	7
6	6	7	6
6	8	6	7
9	10	7	9
9	7	6	7
10	10	10	10
8	8	10	9
7	7	7	7
7	7	7	7
10	8	9	9
8	7	5	7
8	8	6	7
8	8	8	8
10	10	10	10
Scientific correctness Q7 PT	**Comprehensibility Q7 PT**	**Actionability Q7 PT**	**Overall score Q7 PT**
Scientific experts			
4	5	4	4
5	6	5	5
6	5	6	6
5	5	4	5
8	8	9	8
1	4	2	2
6	6	7	6
4	4	4	4
4	3	2	3
PTs			
7	7	8	7
10	10	10	10
7	6	6	6
2	6	3	4
8	6	6	7
7	6	8	7
5	6	5	5
8	7	7	7
8	8	9	8
7	10	7	8
6	7	7	7
7	7	7	7
9	9	9	9
9	9	8	9
2	8	8	6
3	5	5	4
7	8	8	8
5	10	10	8

Scientific correctness Q8 PT	Comprehensibility Q8 PT	Actionability Q8 PT	Overall score Q8 PT
Scientific experts			
2	5	2	3
0	8	8	5
8	8	9	8
7	7	8	7
3	6	4	4
5	6	5	5
5	7	5	6
2	5	4	4
8	10	8	9
PTs			
10	9	7	9
8	8	10	9
6	8	8	7
4	4	0	3
9	7	5	7
6	6	5	6
8	7	7	7
8	8	8	8
5	5	5	5
9	10	9	9
7	7	7	7
7	7	7	7
9	9	9	9
8	9	9	9
5	7	7	6
4	6	5	5
7	7	8	7
7	10	7	8
Scientific correctness Q9 PT	**Comprehensibility Q9 PT**	**Actionability Q9 PT**	**Overall score Q9 PT**
Scientific experts			
3	4	3	3
5	7	8	7
9	9	9	9
9	8	8	8
4	7	5	5
8	8	4	7
7	7	7	7
8	7	8	8
7	9	6	7
PTs			
10	7	8	8
9	9	8	9
6	5	5	5
2	6	2	3
7	9	8	8
8	6	7	7
6	6	4	5
8	10	9	9
8	6	5	6
9	9	9	9
8	8	8	8
7	7	7	7
7	7	7	7
10	10	10	10
9	6	6	7
5	5	6	5
6	5	6	6
0	10	10	7

Table S7. Summary statistics for the grading component "scientific correctness".
Summary Statistics for Scientific Correctness

Question Number		Mean	Median	Interquartile Range	Minimum	Maximum
1	ChatGPT	8.26	8	1.0	5	10
1	PT	5.43	6	4.0	1	10
2	ChatGPT	8.04	8	2.0	4	10
2	PT	5.22	5	2.5	2	10
3	ChatGPT	7.26	8	1.0	1	10
3	PT	6.33	7	3.0	1	10
4	ChatGPT	7.85	8	2.0	6	10
4	PT	7.89	8	2.0	4	10
5	ChatGPT	8.07	8	1.5	2	10
5	PT	7.85	8	2.0	2	10
6	ChatGPT	7.81	8	2.0	5	10
6	PT	7.63	8	2.0	2	10
7	ChatGPT	7.89	8	2.0	4	10
7	PT	5.93	6	3.0	1	10
8	ChatGPT	7.59	8	2.0	4	10
8	PT	6.19	7	3.0	0	10
9	ChatGPT	7.44	8	3.0	3	10
9	PT	6.85	7	2.5	0	10

Table S8. Summary statistics for the grading component "comprehensibility"
Summary Statistics for Comprehensibility

Question Number		Mean	Median	Interquartile Range	Minimum	Maximum
1	ChatGPT	7.67	8	2.0	3	10
1	PT	7.81	8	2.0	4	10
2	ChatGPT	8.63	9	1.5	5	10
2	PT	6.15	6	3.0	1	10
3	ChatGPT	8.48	9	2.0	5	10
3	PT	7.37	8	3.0	3	10
4	ChatGPT	7.81	8	2.0	3	10
4	PT	7.33	7	1.5	3	10
5	ChatGPT	8.44	8	1.5	5	10
5	PT	8.22	8	2.0	6	10
6	ChatGPT	8.37	8	1.0	6	10
6	PT	7.11	7	1.5	3	10
7	ChatGPT	8.52	9	1.0	5	10
7	PT	6.70	6	2.5	3	10
8	ChatGPT	8.26	8	1.0	5	10
8	PT	7.26	7	2.0	4	10
9	ChatGPT	8.04	8	2.0	4	10
9	PT	7.30	7	3.0	4	10

Table S9. Summary statistics for the grading component "actionability"
Summary Statistics for Actionability

Question Number		Mean	Median	Interquartile Range	Minimum	Maximum
1	ChatGPT	6.78	7	4.00	2	10
1	PT	6.83	7	2.50	3	10
2	ChatGPT	7.22	7	3.00	2	10
2	PT	5.11	5	3.50	1	10
3	ChatGPT	8.20	8	1.75	6	10
3	PT	7.33	8	1.00	3	10
4	ChatGPT	6.52	7	1.50	0	10
4	PT	7.11	7	2.00	2	10
5	ChatGPT	7.59	8	1.50	3	10
5	PT	7.26	8	1.00	3	9
6	ChatGPT	7.78	8	2.00	2	10
6	PT	6.48	7	2.50	2	10
7	ChatGPT	8.04	8	1.50	3	10
7	PT	6.44	7	3.00	2	10
8	ChatGPT	7.33	7	3.00	3	10
8	PT	6.52	7	3.00	0	10
9	ChatGPT	7.15	7	3.00	3	10
9	PT	6.78	7	2.50	2	10

Table S10. Sensitivity analysis using the median instead of the mean as the test statistic for group differences between ChatGPT and PT.

	Mean	Median	Mean	Median	Mean	Median	Mean	Median
Question 1	0.01778	0.00395	1e-05	0.03072	0.81334	1	0.934	1
Question 2	<0.0001	0.00465	<0.0001	0.00422	<0.0001	0.06457	1e-05	7e-05
Question 3	0.00041	0.004	0.00356	0.0312	0.00768	0.06548	0.02126	1
Question 4	0.81171	1	1	1	0.16879	1	0.12076	0.18911
Question 5	0.118	0.68951	0.6447	1	0.56767	1	0.31438	1
Question 6	0.00708	0.18943	0.70128	1	0.00106	0.1916	0.01272	0.18976
Question 7	6e-05	0.00097	0.00021	0.03094	0.00013	0.00427	0.00231	0.01247
Question 8	0.00474	0.00417	0.02595	0.28711	0.00388	0.00416	0.03549	0.40283
Question 9	0.0535	0.18872	0.30964	1	0.04379	0.18761	0.30654	0.54254

Table S11. Flesch Reading Ease (FRE) score and the Flesch–Kincaid Grade Level analysis

ChatGPT_FRE	ChatGPT_FKGL	PT_FRE	PT_FKGL
39.16236565	11.88473129	61.9276667	9.9877037
46.46332501	10.36121783	51.175	11.09
53.68649827	9.699404644	64.2640385	9.14012821
41.38750055	11.4176077	60.0575	9.23
51.20650273	11.1684153	38.1017241	15.1493103
44.70138251	11.06547541	12.7697627	16.3846519
57.30355769	10.15875	47.5343521	13.6845915
40.14042135	12.39841894	48.89	10.415

Table S12. Flesch Reading Ease (FRE) score and the Flesch–Kincaid Grade Level summary statistics

	ChatGPT_FRE	ChatGPT_FKGL	PT_FRE	PT_FKGL
count	8	8	8	8
mean	46.75644422	11.01925264	48.0900055	11.8851732
std	6.69377978	0.906303254	16.7077928	2.80604254
min	39.16236565	9.699404644	12.7697627	9.14012821
25%	41.07573075	10.31060087	45.1761951	9.79827778
50%	45.58235376	11.11694536	50.0325	10.7525
75%	51.82650162	11.5343886	60.5250417	14.0507712
max	57.30355769	12.39841894	64.2640385	16.3846519

Table S13. Intraclass correlation coefficient (ICC) among grading PTs

criterion	Answer set	ICC (2,k)	lower bound CI	upper bound CI	nr of Qs	nr of graders	Inter-pretation
Actionability	ChatGPT	0.69108278	0.390892658	0.906857013	9	18	Good
Actionability	PT	0.61026517	0.24603212	0.881856518	9	18	Good
Comprehen-sibility	ChatGPT	0.634818	0.289576003	0.889575215	9	18	Good
Comprehen-sibility	PT	0.59796979	0.200445836	0.881201226	9	18	Fair
Scientific correctness	ChatGPT	0.27672621	-0.220600161	0.754692695	9	18	Poor
Scientific correctness	PT	0.76384081	0.49759271	0.933323246	9	18	Excellent

Table S14. Intraclass correlation coefficient (ICC) among scientific experts grading the five questions relating to the topic "training for fat loss"

criterion	Answer set	ICC (2,k)	lower bound CI	upper bound CI	nr of Qs	nr of graders	Inter-pretation
Actionability	ChatGPT	-0.116451	-0.330532978	0.56717331	5	9	Poor
Actionability	PT	0.54667368	0.046674825	0.930112294	5	9	Moderate
Comprehen-sibility	ChatGPT	-0.0503919	-0.371801441	0.689715167	5	9	Poor
Comprehen-sibility	PT	0.77114428	0.38094296	0.970030924	5	9	Good
Scientific correctness	ChatGPT	0.2901169	-0.270922221	0.875311571	5	9	Poor
Scientific correctness	PT	0.84285411	0.540023571	0.980256013	5	9	Good

Table S15. Mean difference of standard deviation (SD) between ChatGPT and PTs per question

topic	question	criterion	mean difference SD between ChatGPT and PTs
timing of the day for training	Q7	Actionability	-0.131782601
timing of the day for training	Q7	Comprehensibility	0.128172516
timing of the day for training	Q7	Scientific correctness	-0.804059777
training around pain	Q6	Actionability	-1.013699052
training around pain	Q6	Comprehensibility	-1.061552813
training around pain	Q6	Scientific correctness	-0.334237951
training for fat loss	Q2	Actionability	0.19584231
training for fat loss	Q2	Comprehensibility	-0.256565336
training for fat loss	Q2	Scientific correctness	-0.01766691
training for fat loss	Q4	Actionability	0.052707882
training for fat loss	Q4	Comprehensibility	-0.115893286
training for fat loss	Q4	Scientific correctness	-0.242666042
training for fat loss	Q5	Actionability	0.638118931
training for fat loss	Q5	Comprehensibility	0.356393959
training for fat loss	Q5	Scientific correctness	0.215938357
training for fat loss	Q8	Actionability	-0.105565843
training for fat loss	Q8	Comprehensibility	-0.034754456
training for fat loss	Q8	Scientific correctness	-1.462899188
training for fat loss	Q9	Actionability	0.502929162
training for fat loss	Q9	Comprehensibility	0.5
training for fat loss	Q9	Scientific correctness	-0.058129128
training frequency	Q1	Actionability	-0.144918045
training frequency	Q1	Comprehensibility	-1.572025607
training frequency	Q1	Scientific correctness	-1.578524259
training motivation	Q3	Actionability	-1.407179419
training motivation	Q3	Comprehensibility	-0.4507505
training motivation	Q3	Scientific correctness	-0.72625733

Table S16. Mean difference of standard deviation (SD) between ChatGPT and PTs across all nine questions

criterion	mean difference SD between ChatGPT and PTs across all 9 questions
Actionability	-0.157060742
Comprehensibility	-0.278552836
Scientific correctness	-0.556500248