Biomedical Human Kinetics
Volume 18 Issue 1 · pp. 337–345 · DOI: 10.2478/bhk-2026-0033

Biomedical Human Kinetics

Evaluation of ChatGPT-generated training programmes for kickboxing

Research Article

Denis Volsky[1]*, Georgiy Korobeynikov[2,3], Olha Borysova[4], Inna Holovach[5]

[1] *National Federation of Kickboxing of Ukraine WAKO, Ukraine*
[2] *Department of Theory and Methods of International Wrestling, Uzbek State University of Physical Education and Sport, Chirchik, Uzbekistan*
[3] *Department of Combat Sports and Power Sports, National University of Ukraine on Physical Education and Sport, Kyiv, Ukraine*
[4] *Department of Professional, Non-Olympic and Adaptive Sports, National University of Ukraine on Physical Education and Sport, Kyiv, Ukraine*
[5] *Department of Physical Education and Pedagogy of Sport, Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine*

Received 06 May 2026; Accepted 25 September 2026

Abstract

Study aim: The aim of this research is to evaluate the expert-perceived quality of ChatGPT-generated training programmes for kickboxing.
Material and methods: Three 8-week training programmes for beginner-level kickboxers were generated using prompts with increasing levels of contextual detail. Eleven experienced kickboxing coaches evaluated the programmes using a structured questionnaire consisting of 12 items related to programme organization, sport-specific preparation, and practical applicability. Each item was rated using a five-point Likert scale. Descriptive statistics were calculated, and differences between the programmes were analysed using the Friedman test with *post hoc* Wilcoxon signed-rank comparisons. Internal consistency and inter-expert agreement were assessed using Cronbach's α and Kendall's coefficient of concordance.
Results: The expert-perceived quality of the programmes improved with increasing prompt specificity. The mean composite expert score increased from 1.33 ± 0.47 for Training Programme 1 (Basic prompt) to 2.25 ± 0.89 for Training Programme 3 (Detailed prompt). Statistically significant differences were observed between all program. The largest improvements were observed in programme organization, training progression, training diversity, and physical conditioning. Internal consistency was high across all evaluations (Cronbach's α = 0.85–0.92).
Conclusion: The findings indicate that the quality of AI-generated training programmes strongly depends on the specificity of the prompts used to generate them. When detailed contextual information is provided, ChatGPT may serve as a supportive tool for generating preliminary training structures. However, the generated programmes should be regarded as preliminary drafts requiring expert review to ensure methodological accuracy, sport-specific relevance, and safe implementation in coaching practice.

Keywords: *Combat sport • Martial arts • Psychophysiology • Sport psychology • Sport physiology*

1. Introduction

In contemporary sport, technologies are playing an increasingly significant role in athlete preparation, training management, and physiological monitoring. Digital technological solutions are continually transforming approaches to training, recovery, and performance assessment (Curby et al., 2023; Ráthonyi et al., 2018). Such technologies range from sensor devices and tracking systems to analytical platforms and virtual simulators. These tools enable coaches and researchers to exercise more precise control over training loads, individualize preparation programmes, and make decisions based on objective data. These developments facilitate a

Author's address: Denis Volsky, National Federation of Kickboxing of Ukraine WAKO, Ukraine athletefc@gmail.com

Open Access. © 2026 the author(s), published by University of Physical Education in Warsaw
This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

more objective and scientific approach, involving the use of integrated digital ecosystems to support athletes throughout all stages of preparation (Toto, 2025). In this context, the coach acts not only as a physical training specialist but also as a user of technological systems, necessitating a new level of digital competence, data analytics knowledge, and understanding of the potential of contemporary technologies in sport (Mănescu & Mănescu, 2025).

In recent years, artificial intelligence (AI) has been rapidly advancing and finding applications across various domains, including medicine, education, manufacturing, and sport. Significant progress in machine learning, computer vision, and natural language processing has created new opportunities for analysing large datasets, predicting outcomes, and generating personalized solutions (Bucea-Manea-Țoniș et al., 2025). Within sports science, research has been conducted to address a wide range of tasks using AI (Muñoz-Macho et al., 2024), such as competition outcome prediction, injury prevention algorithms, and monitoring and controlling athlete training loads. Particular attention has been given to computer vision technologies, notably pose estimation techniques, for analysing the biomechanics and technical execution of movements (Latyshev et al., 2024). These advancements open up promising prospects for employing AI as a tool to enhance the efficiency of athlete preparation systems.

A particularly important area of AI development is large language models (LLMs), whose popularity has grown rapidly in recent years. They have demonstrated significant potential in text generation, providing consultations, and automating routine tasks in scientific and educational contexts (Lai et al., 2025; Saraç et al., 2025). In the field of sport, the use of LLMs remains in its early stages, yet their adoption is rapidly emerging, with a growing number of studies beginning to investigate their potential applications. Researchers have investigated various uses of LLMs in sport (An, 2025; Qiu, 2024; Ha et al. 2024) analysed the potential of ChatGPT in kinesiology, particularly for designing training tasks and exercise scenarios; the potential of LLMs has also been demonstrated for predicting tactical decisions in football and modelling team strategic behaviour (Caron & Müller, 2023); additionally, the use of LLMs for interpreting information in sports medicine has been examined (Shan et al., 2025).

A particularly relevant application of LLMs lies in the development of individualized training programmes for both professional athletes and individuals engaging in physical activity for health maintenance. Contemporary research highlights the active integration of AI technologies in the design of such programmes. Specifically, Düking et al. (2024) evaluated ChatGPT-generated training plans for runners with varying levels of input detail; Wachholz et al. (2025) investigated the level of trust in AI-generated plans among recreational athletes and assessed their quality as rated by experienced coaches; Havers et al. (2025) analysed the quality and reproducibility of strength-training programmes created by GPT-4 and Google Gemini; while Puce et al. (2025), in a systematic review, examined the use of LLM tools for developing training and wellness recommendations and assessed their effectiveness across different contexts. Furthermore, recent studies emphasize the expanding capabilities of LLMs in providing personalized guidance on nutrition, recovery strategies, and even in the development of educational projects within sports science (Van Arum et al., 2025). Recent research has also explored AI-powered digital coaching systems as potential decision-support tools for coaches, further expanding the practical applications of LLMs in sport (Genç et al., 2025). Among all LLMs, ChatGPT stands out as the most widely adopted and frequently utilized model in sports science research. Nevertheless, questions regarding the reliability, consistency, and quality of generated programmes – particularly in specific sports disciplines – remain open and warrant systematic scientific evaluation. Although previous studies have evaluated AI-generated training programmes in endurance and strength-related activities, little is known about their applicability in combat sports. Kickboxing represents a particularly demanding context because training programmes must integrate technical, tactical, physical, and psychological components while ensuring appropriate progression and safety for novice athletes. Consequently, it remains unclear whether current LLMs can generate programmes that adequately reflect these sport-specific requirements. Future research in this area will enable the assessment of the practical potential of LLMs in physical education and sport, as well as identify promising pathways for their integration into the training process.

The aim of this research is to evaluate the expert-perceived quality of ChatGPT-generated training programmes for kickboxing. It was hypothesized that increasing the specificity of prompts provided to ChatGPT would result in higher-quality training programmes, as evaluated by expert coaches.

2. Material and methods

2.1 Participants

A total of 11 expert coaches participated in the expert evaluation. Each expert was required to have at least

5 years of coaching experience in kickboxing and to have trained a minimum of two qualified athletes in accordance with the current national sport classification system corresponding to Level 3 (McKay et al., 2021). The participating expert coaches had a mean coaching experience of 13.27 ± 6.48 years and a mean age of 40.18 ± 11.67 years and were actively coaching athletes competing at regional and national levels at the time of the study. All expert coaches had practical experience in planning and implementing annual training programmes for beginner and competitive kickboxers. All experts voluntarily agreed to participate in the study and provided written informed consent prior to the evaluation. The study protocol was approved by the institutional ethics committee of Department of Combat Sports and Power Sports, National University of Ukraine on Physical Education and Sport (Protocol No. 1, August 28, 2025) and conducted in accordance with the principles outlined in the Declaration of Helsinki.

2.2 Expert questionnaire

The questionnaire was adapted from previously published evaluation frameworks used for assessing AI-generated training programmes (Düking et al., 2024). Additional adaptations were made to reflect the specific methodological and technical characteristics of kickboxing training (Toros, 2011). The selection and wording of the questionnaire items were informed by previous studies evaluating ChatGPT-generated training programmes (Ayers et al., 2023; Shan et al., 2025; Wachholz et al., 2025) and related applications in medicine (Lukac et al., 2025).

The final questionnaire consisted of 12 items assessing different aspects of training programme design. The items were grouped into three major domains (Table 1): overall structure and methodological principles of the training program; combat sports specificity; and practicality and applicability. Each item was rated using a five-point Likert scale (1–5), where lower scores indicated poorer programme quality and higher scores reflected better compliance with methodological principles of training design. The scale descriptors were as follows: 1 = very poor, 2 = poor, 3 = acceptable, 4 = good, and 5 = excellent.

2.3 Training programme generation procedure

To generate the training programmes using ChatGPT, the basic preparation stage of athletes (first and second year of training) was selected, as it represents a critical period for developing fundamental motor skills and general physical fitness and provides a suitable context for evaluating the quality of automatically generated training plans (Bompa & Buzzichelli, 2019). The generated training programmes were designed within the context of WAKO (World Association of Kickboxing Organizations) kickboxing rules and training principles. To consistently refine the prompts and assess the impact of contextual detail on the quality of generated programmes, three different text instructions (prompts) were used. Each prompt included progressively more detailed information regarding the content, objectives, and structure of the training process. This approach aligns with previous research emphasizing the importance of prompt clarity when interacting with LLMs to obtain more scientifically grounded outcomes (Wachholz et al., 2025; Yang et al., 2025).

The following prompts were used:

1. Training programme 1 (TP1, Basic prompt): Create an 8-week training programme for a group of beginner-level kickboxing athletes.
2. Training programme 2 (TP2, Refined prompt): Create an 8-week training programme for a group of beginner-level kickboxing athletes. Consider that the athletes have no prior kickboxing experience (or up to 1 year) but may have participated in other sports. The goal of the programme is general physical development and mastery of basic technical elements (stance, movement, simple punches).
3. Training programme 3 (TP3, Detailed prompt): Create an 8-week training programme for a group of beginner-level kickboxing athletes. The athletes are 10–11 years old, have no medical contraindications for sports participation, have undergone a medical examination, and are cleared for training. The group includes several weight categories. The participants have no prior kickboxing experience (or up to 1 year) but may have backgrounds in other sports, including combat disciplines. The total duration of the programme is 8 weeks, with the first week primarily focused on orientation and adaptation. The main objectives of the programme are general physical development, acquisition of fundamental technical elements (stance, footwork, straight punches, and kicks), and improvement of motor coordination. Training sessions should be conducted three times per week with progressive load increases. The programme should include a warm-up, main part, and cool-down, as well as an alternation of general and specific physical preparation, taking into account safety requirements and the availability of basic equipment (punching bag, pads, and elementary protective gear).

Table 1. Proposed questionnaire items and key constructs for training programme evaluation

Aspect	Questionnaire items	Key construct
Overall structure and methodological principles	Rate the overall structure of the program	Program
	Rate the progression of intensity and/or volume	Progression
	Evaluate the extent to which the programme considers athletes' preparedness	Individualization
	Assess the inclusion of different preparation components (technical, physical, tactical, etc.)	Diversity
	Rate the integration of injury prevention, rest, and recovery measures	Safety
Combat sports specificity	Rate the exercises aimed at developing technical skills (strikes, stance, footwork)	Technical
	Evaluate the extent to which the programme develops tactical skills (combinations, decision-making, strategy)	Tactical
	Rate the coverage of main physical qualities (strength, endurance, speed, agility)	Conditioning
	Assess the inclusion of psychological skills development	Psychological
	Rate the systematic use of partner work (sparring, pair drills)	Sparring
Practicality and applicability	Rate the feasibility of implementation under typical training conditions (gym, equipment, time)	Feasibility
	Evaluate the suitability of the programme for beginners in group training	Novices

Source: Authors' contribution.

The expert coaches were informed that all training programmes had been generated using ChatGPT, as the primary aim of the study was to evaluate the influence of prompt specificity on AI-generated outputs. However, the programmes were presented in a randomized order, and the experts were not informed which programme corresponded to each prompt or level of prompt specificity.

Training programmes were generated using ChatGPT (OpenAI, GPT-4o) through the publicly available free web interface on November 14, 2025. The default system settings were used throughout programme generation. No additional plugins, external software tools, or custom GPTs were employed. All training programmes were generated directly by the authors without assistance from third-party software or external editing tools. The AI-generated training programmes were presented to the expert coaches without any modification to their content, ensuring that the evaluations reflected the original ChatGPT outputs. No generative AI tools were used for interpretation or data analysis; AI use was limited solely to language editing support.

2.4 Statistical analysis

For each training program, descriptive statistics were calculated, including the mean value, standard deviation (SD), median, and interquartile range (IQR) for all evaluation items on the Likert scale. Composite expert scores were also calculated for each program, with the calculation of internal consistency and inter-rater reliability coefficients. Cronbach's α coefficient was used to assess the internal consistency of the scale, while Kendall's coefficient of concordance (W) was applied to determine the level of inter-expert agreement. To examine differences between the three training programmes, a Friedman test for repeated measures was first applied. When significant differences were detected, pairwise comparisons were conducted using the Wilcoxon signed-rank test with Bonferroni correction for multiple comparisons. All calculations and visualizations were conducted in Python (version 3.10) using the pandas, numPy, sciPy, and matplotlib libraries (Virtanen et al., 2020).

3. Results

The results of the expert evaluation of the three ChatGPT-generated training programmes are presented in Table 2. Descriptive statistics summarize experts' ratings across 12 key constructs reflecting programme quality, structure, and applicability.

The overall expert rating increased progressively across the programmes. The mean composite score increased from 1.33 ± 0.47 for TP1 to 1.78 ± 0.69 for TP2

Table 2. Comparative descriptive statistics of expert assessments for each training program

Key construct	TP1, basic prompt		TP2, refined prompt		TP3, detailed prompt	
	Mean value ± SD	Median (IQR)	Mean value ± SD	Median (IQR)	Mean value ± SD	Median (IQR)
Programme	1.36 ± 0.50	1 (1–2)	1.91 ± 0.70	2 (2–2)	2.82 ± 0.75	3 (2–3)
Progression	1.27 ± 0.47	1 (1–2)	1.82 ± 0.75	2 (1–2)	3.00 ± 0.77	3 (2–4)
Individualization	1.18 ± 0.40	1 (1–1)	1.55 ± 0.69	1 (1–2)	2.09 ± 0.94	2 (2–2)
Diversity	1.45 ± 0.52	1 (1–2)	1.91 ± 0.70	2 (2–2)	2.82 ± 0.75	3 (2–3)
Safety	1.18 ± 0.40	1 (1–1)	1.36 ± .050	1 (1–2)	1.64 ± 0.67	2 (1–2)
Technical	1.36 ± 0.50	1 (1–2)	2.45 ± 0.52	2 (2–3)	2.00 ± 0.77	2 (2–2)
Tactical	1.18 ± 0.40	1 (1–1)	1.64 ± 0.67	2 (1–2)	1.73 ± 0.65	2 (1–2)
Conditioning	1.64 ± 0.50	2 (1–2)	2.00 ± 0.77	2 (2–2)	2.64 ± 0.92	3 (2–3)
Psychological	1.18 ± 0.40	1 (1–1)	1.45 ± 0.52	1 (1–2)	1.73 ± 0.65	2 (1–2)
Sparring	1.18 ± 0.40	1 (1–1)	1.45 ± 0.52	1 (1–2)	1.91 ± 0.70	2 (2–2)
Feasibility	1.55 ± 0.52	2 (1–2)	1.64 ± 0.50	2 (1–2)	2.82 ± 0.75	3 (2–3)
Novices	1.36 ± 0.50	1 (1–2)	2.18 ± 0.75	2 (2–3)	1.82 ± 0.75	2 (1–2)
Overall expert rating	1.33 ± .47	1 (1–2)	1.78 ± 0.69	2 (1–2)	2.25 ± 0.89	2 (2–3)

Source: Authors' contribution.

and 2.25 ± 0.89 for TP3. Median values also increased from 1 (1–2) for TP1 to 2 (1–2) for TP2 and 2 (2–3) for TP3.

Across individual constructs, TP3 received higher ratings than TP1 for programme organization (2.82 and 1.36 points, respectively), training progression (3.00 and 1.27 points), training diversity (2.82 and 1.45 points), and physical conditioning (2.64 and 1.64 points). Smaller differences between the programmes were observed for safety and recovery, psychological preparation, and tactical preparation. In two constructs, technical preparation and applicability for novices, TP2 received slightly higher ratings than TP3.

The internal consistency of the expert ratings was high for all programme evaluations. Cronbach's α coefficients were 0.92 for TP1, 0.88 for TP2, and 0.85 for TP3. Inter-expert agreement assessed using Kendall's coefficient of concordance was 0.19 ($p < 0.05$) for TP1, 0.32 ($p < 0.01$) for TP2, and 0.48 ($p < 0.01$) for TP3.

The distribution of composite expert ratings for each programme is presented in Figure 1. The boxplots illustrate a progressive shift toward higher ratings from TP1 to TP3.

The Friedman test indicated statistically significant differences between the three programmes ($\chi^2 = 18.74$, $p < 0.01$). Pairwise comparisons using the Wilcoxon signed-rank test with Bonferroni adjustment showed that TP2 received higher ratings than TP1 ($p < 0.01$), TP3 received higher ratings than TP2 ($p < 0.01$), and TP3 also received higher ratings than TP1 ($p < 0.01$).

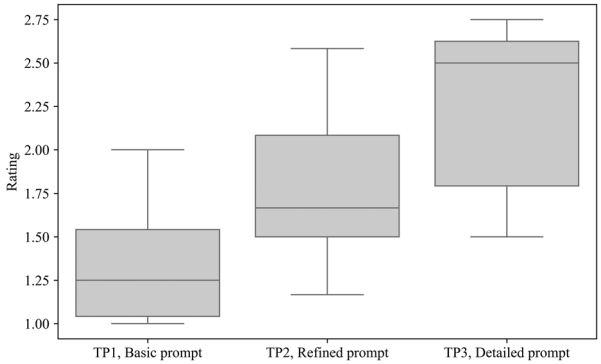


Figure 1. Boxplot of composite expert ratings for each training program

Source: Authors' contribution.

4. Discussion

In contemporary sports settings, the application of AI is becoming increasingly widespread, offering new opportunities for the planning of training programmes and the optimization of athlete preparation. Tools based on LLMs, such as ChatGPT, enable coaches and athletes to rapidly generate structured training programmes that adhere to specific criteria for intensity, progression, and sport-specific requirements. However, the effectiveness of these programmes largely depends on the quality of input data and the clarity of instructions provided to the AI (Caron & Müller, 2023; An, 2025).

In the present study, ChatGPT was employed to create three distinct athlete training programmes (TP1–TP3), which were subsequently evaluated by a panel of experts across 12 key components. The analysis revealed variable quality among the generated programmes. Cronbach's α values ranged from 0.85 to 0.92, indicating high internal consistency of the ratings. Meanwhile, Kendall's W varied from 0.19 ($p < 0.05$) for TP1 to 0.48 ($p < 0.01$) for TP3, indicating differences in the level of inter-expert concordance. Although Kendall's W increased progressively across the three programmes, the observed coefficients indicate only weak to moderate agreement among experts. Therefore, despite statistically significant differences between the programmes, the findings should be interpreted with appropriate caution, as expert consensus remained limited. A similar trend was observed in the study by Düking et al. (2024), where Fleiss' Kappa values for three comparable programmes were 0.43, 0.25, and 0.00 (all $p < 0.01$), supporting the general pattern of decreasing agreement with increasing programme complexity or ambiguity.

A detailed analysis of the programme components highlights the blocks that received the highest ratings from the experts: training process organization, training progression and exercise diversity, safety, and physical conditioning. In contrast, the components related to psychological and tactical preparation, as well as applicability for novices, demonstrated lower agreement in certain programmes, suggesting potential challenges in formulating universal recommendations for these aspects (Ha et al., 2024; Saraç et al., 2025; Toros, 2011).

Our findings further confirm that clear and detailed instructions provided to ChatGPT significantly affect the quality of the generated training programmes. Increasing the level of prompt specificity enables the model to produce more structured and balanced programmes. Specifically, the average quality ratings increased from 1.33 for the basic prompt (TP1) to 1.78 for the refined prompt (TP2), and up to 2.25 for the detailed prompt (TP3). This trend aligns with expert assessments reported in other studies (Düking et al., 2024), emphasizing the importance of precise prompt formulation to achieve optimal quality of AI-generated content. These results underscore the value of interactive human–AI collaboration, where expert oversight ensures that the outputs meet the actual needs of athletes. Nevertheless, despite the statistically significant improvement across TP1–TP3, the absolute expert ratings remained relatively low, with the highest-rated programme (TP3) achieving a mean score of only 2.25 out of 5. Therefore, the observed improvements should be interpreted as relative rather than indicating that the generated programmes reached a level suitable for direct implementation in coaching practice.

At the same time, the application of AI for generating training programmes has certain limitations. The relatively low to moderate Kendall's coefficients observed in the present study suggest that expert agreement remained limited despite the improvement in programme quality. Similar observations regarding variability in expert agreement have been reported in previous studies evaluating AI-generated training programmes (Düking et al., 2024). Also, this observation is consistent with previous research, which suggests that LLMs cannot always provide optimal levels of detail and accuracy without supplementary input data (Ayers et al., 2023; Shan et al., 2025). Therefore, AI serves as an effective supportive tool but cannot fully replace expert evaluation and supervision.

From a practical perspective, the present findings suggest that increasing prompt specificity improves the structure and overall quality of AI-generated training programmes. However, all programmes required expert refinement before practical application. Consequently, ChatGPT-generated programmes should be regarded as preliminary drafts that may support coaches during the planning process rather than as ready-to-use training programmes. These results may be relevant across a wide range of sports and support the development of hybrid planning models that integrate AI with expert oversight (Van Arum et al., 2025; Wachholz et al., 2025; Yang et al., 2025).

An additional consideration concerns the level of trust in such technologies. In the study by Wachholz et al. (2025), participants using AI-based tools exhibited higher levels of trust in the technology ($p < 0.05$) compared with non-users, highlighting the importance of coaches' technological competence. Overall, the study demonstrates the potential of AI in sports training while simultaneously revealing its limitations and the necessity for expert intervention to ensure programme accuracy and effectiveness. Future

research may focus on evaluating AI-generated training programmes for athletes of varying skill levels, integrating individual athlete data, testing programmes in real-world training conditions, and comparing AI-generated programmes with recognized coaching frameworks (e.g., WAKO-based recommendations) to provide a more objective benchmark for their evaluation.

Several limitations of the present study should be acknowledged. First, the number of expert coaches participating in the evaluation was relatively small, which may limit the generalizability of the findings. Second, the assessment relied on subjective expert judgments rather than objective measures of athlete performance. Although the questionnaire was adapted from previously published evaluation frameworks, the adapted version was not subjected to a separate formal validation procedure, which should be addressed in future research. Third, the participating experts were aware that all evaluated training programmes had been generated by ChatGPT. Although the programmes were presented in a randomized order and their prompt origin was concealed, prior attitudes toward AI-assisted coaching may have influenced expert judgments, introducing potential expectation bias. Furthermore, the relatively low to moderate Kendall's coefficients indicate that expert agreement was not uniform across all programme evaluations and should therefore be considered when interpreting the findings. Finally, the AI-generated programmes were not implemented or evaluated under real training conditions; consequently, their practical effectiveness and long-term impact on athlete development remain uncertain. The study also focused exclusively on beginner-level WAKO kickboxing athletes, which may limit the generalizability of the findings to other combat sports, rule sets, or athlete populations with different levels of expertise.

5. Conclusion

The results of the study confirmed that the quality of training programmes generated using ChatGPT largely depends on the level of detail provided in the prompt. Gradual refinement of the instructions led to an increase in mean expert ratings from 1.33 ± 0.47 (TP1) to 2.25 ± 0.89 (TP3), indicating improvements in programme structure, balance, and practical applicability. High levels of internal consistency (Cronbach's α = 0.85–0.92) support the reliability of the ratings, whereas variations in inter-rater agreement (Kendall's W = 0.19–0.48) highlight subjective differences in expert perception. Although more detailed prompts improved the quality of the generated programmes, the expert ratings indicate that these programmes should be regarded as preliminary drafts rather than training plans ready for direct implementation in coaching practice. Therefore, ChatGPT may serve as a supportive tool for generating initial training programme concepts, while expert review and refinement remain essential to ensure programme quality, safety, and individualization.

Acknowledgments

The authors express their sincere gratitude to all experts and participants who voluntarily took part in this study.

Funding information

This research received no external funding.

Author contributions

Denis Volsky and Georgiy Korobeynikov contributed to the conceptualization and design of the study. Inna Holovach contributed to data collection, data processing, and statistical analysis. Olha Borysova contributed to the interpretation of the results, formulation of the conclusions, and critical revision of the manuscript. Denis Volsky and Georgiy Korobeynikov contributed to drafting and revising the manuscript. All authors read and approved the final version of the manuscript.

Conflict of interest statement

The authors declare no conflict of interest.

References

An, R. (2025). Artificial intelligence in health and sport sciences: Promise, progress, and prudence. *Journal of Sport and Health Science*, *14*, 101054. doi: 10.1016/j.jshs.2025.101054

Ayers, J. W., Poliak, A., Dredze, M., Leas, E. C., Zhu, Z., Kelley, J. B., et al. (2023). Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum.

JAMA Internal Medicine, *183*(6), 589–596. doi: 10.1001/jamainternmed.2023.1838

Bompa, T. O., & Buzzichelli, C. (2019). *Periodization: Theory and methodology of training* (6th ed.). Human Kinetics.

Bucea-Manea-Țoniș, R., Vasile, L., Trușcă, A., & Stănescu, M. (2025). Harnessing artificial intelligence in sports training: Evidence from Romanian professionals using SEM analysis. *Applied Sciences*, *15*(16), 8853. doi: 10.3390/app15168853

Caron, M., & Müller, O. (2023). TacticalGPT: Uncovering the potential of LLMs for predicting tactical decisions in professional football. In StatsBomb Conference Proceedings (pp. 1–11).

Curby, D., Dokmanac, M., Kerimov, F., Tropin, Y., Latyshev, M., Bezkorovainyi, D., et al. (2023). Performance of wrestlers at the Olympic Games: Gender aspect. *Pedagogy of Physical Culture and Sports*, *27*(6), 487–493. doi: 10.15561/26649837.2023.0607

Düking, P., Sperlich, B., Voigt, L., Van Hooren, B., Zanini, M., & Zinner, C. (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 & Medicine*, *23*(1), 56–64. doi: 10.52082/jssm.2024.56

Genç, A., Aydin, R., Canuzakov, K., Dzhiparkul, A., Gürcan, H. H., Dere, G., et al. (2025). Digital coaches: an alternative to expert coaches for men's fitness goals. *Physical Activity Review*, *13*(2), 35. doi: 10.16926/par.2025.13.18

Ha, T., Yu, H., Lim, B., Jung, H. Y., Brooks, C., Krause, J., et al. (2024). Using ChatGPT in the field of kinesiology: Opportunities and considerations. *Journal of Physical Education and Sport*, *24*(1), 3–12. doi: 10.7752/jpes.2024.01001

Havers, T., Masur, L., Isenmann, E., Geisler, S., Zinner, C., Sperlich, B., et al. (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. doi: 10.5114/biolsport.2025.145911

Lai, X., Chen, J., Lai, Y., Huang, S., Cai, Y., Sun, Z., et al. (2025). Using large language models to enhance exercise recommendations and physical activity in clinical and healthy populations: A scoping review. *JMIR Medical Informatics*, *13*(1), e59309. doi: 10.2196/59309

Latyshev, M., Lopatenko, G., Shandryhos, V., Yarmoliuk, O., Pryimak, M., & Kvasnytsia, I. (2024). Computer vision technologies for human pose estimation in exercise: Accuracy and practicality. In Society. Integration. Education. Proceedings of the International Scientific Conference (Vol. 2, pp. 626–636). doi: 10.17770/sie2024vol2.7842

Lukac, P. J., Turner, W., Vangala, S., Chin, A. T., Khalili, J., Shih, Y. C. T., et al. (2025). A randomized clinical trial of two ambient artificial intelligence scribes: Measuring documentation efficiency and physician burnout [Preprint]. medRxiv. doi: 10.1101/2025.07.10.25331333

Mănescu, D. C., & Mănescu, A. M. (2025). Artificial intelligence in the selection of top-performing athletes for team sports: A proof-of-concept predictive modeling study. *Applied Sciences*, *15*(18), 9918. doi: 10.3390/app15189918

McKay, A. K., Stellingwerff, T., Smith, E. S., Martin, D.T., Mujika, I., Goosey-Tolfrey, V. L., et al. (2021). Defining training and performance caliber: A participant classification framework. *International Journal of Sports Physiology and Performance*, *17*(2), 317–331. doi: 10.1123/ijspp.2021-0451

Muñoz-Macho, A. A., Domínguez-Morales, M. J., & Sevillano-Ramos, J. L. (2024). Performance and healthcare analysis in elite sports teams using artificial intelligence: A scoping review. *Frontiers in Sports and Active Living*, *6*, 1383723. doi: 10.3389/fspor.2024.1383723

Puce, L., Bragazzi, N. L., Curra, A., & Trompetto, C. (2025). Harnessing generative artificial intelligence for exercise and training prescription: Applications and implications in sports and physical activity–A systematic literature review. *Applied Sciences*, *15*(7), 3124. doi: 10.3390/app15073497

Qiu, Y. (2024). The impact of LLM hallucinations on motor skill learning: A case study in badminton. *IEEE Access*, *12*, 139669–139682. doi: 10.1109/ACCESS.2024.3444783

Ráthonyi, G., Müller, A., & Rathonyi-Odor, K. (2018). How digital technologies are changing sport? *APSTRACT: Applied Studies in Agribusiness and Commerce*, *12*, 89–96. doi: APSTRACT/2018/3-4/10">10.19041/APSTRACT/2018/3-4/10

Saraç, H., Ulusoy, İ. T., Alpay, J., Ödemiş, H., & Söğüt, M. (2025). Evaluating the potential role of AI chatbots in designing personalized exercise programs for weight management. *International Journal of Human–Computer Interaction*, 1–8. doi: 10.1080/10447318.2025.2462752

Shan, G., Chen, X., Wang, C., Liu, L., Gu, Y., Jiang, H., et al. (2025). Comparing diagnostic accuracy of clinical professionals and large language models: Systematic review and meta-analysis. *JMIR Medical Informatics*, *13*(1), e64963. doi: 10.2196/64963

Toros, T. (2011). Training exercise performance questionnaire (TEPQ) – Development study: A study on

sportsmen from branches of judo, taekwondo, karate. *Archives of Budo*, *7*(2), 81–86.

Toto, G. A. (2025). A systematic review on digital technologies in sport science: Didactics of sport. *Formazione & Insegnamento*, *20*(1 Suppl), 1–12. doi: 10.7346/-feis-XX-01-22_01

Van Arum, S., Genç, H. U., Reidsma, D., & Karahanoğlu, A. (2025). Selective trust: Understanding human-AI partnerships in personal health decision-making process. In Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (pp. 1–21). doi: 10.1145/3706598.3713462

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. doi: 10.1038/s41592-019-0686-2

Wachholz, F., Manno, S., Schlachter, D., Gamper, N., & Schnitzer, M. (2025). Acceptance and trust in AI-generated exercise plans among recreational athletes and quality evaluation by experienced coaches: A pilot study. *BMC Research Notes*, *18*(1), 112. doi: 10.1186/s13104-025-07172-9

Yang, J., Qin, S., & Ren, D. (2025). Artificial intelligence coaches on the sidelines: Evaluating readability and quality of soccer training plans from six generative models. *International Journal of Sports Science & Coaching*, Advance online publication, *21*(1), 75–81. doi: 10.1177/17479541251369593