

Systematic Review

Harnessing Generative Artificial Intelligence for Exercise and Training Prescription: Applications and Implications in Sports and Physical Activity—A Systematic Literature Review

Luca Puce ^{1,2} , Nicola Luigi Bragazzi ^{1,2,*} , Antonio Currà ³  and Carlo Trompetto ^{1,2} 

¹ Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOEMI), University of Genoa, 16132 Genoa, Italy; luca1puce@gmail.com (L.P.); ctrompetto@neurologia.unige.it (C.T.)

² IRCCS Ospedale Policlinico San Martino, 16132 Genoa, Italy

³ Academic Neurology Unit, A. Fiorini Hospital, Terracina, LT, Department of Medico-Surgical Sciences and Biotechnologies, “Sapienza” University of Rome, 00185 Rome, Italy; antonio.curra@uniroma1.it

* Correspondence: robertobragazzi@gmail.com

Abstract: Regular physical activity plays a critical role in health promotion and athletic performance, necessitating personalized exercise and training prescriptions. While traditional methods rely on expert assessments, artificial intelligence (AI), particularly generative AI models such as ChatGPT and Google Gemini, has emerged as a potential tool for enhancing personalization and scalability in training recommendations. However, the applicability, reliability, and adaptability of AI-generated exercise prescriptions remain underexplored. A comprehensive search was performed using the UnoPerTutto meta-database, identifying 2891 records. After duplicate removal (1619 records) and screening, 61 full-text reports were assessed for eligibility, resulting in the inclusion of 10 studies. The studies varied in methodology, including qualitative assessments, mixed-methods approaches, quasi-experimental designs, and a randomized controlled trial (RCT). AI models such as ChatGPT-4, ChatGPT-3.5, and Google Gemini were evaluated across different contexts, including strength training, rehabilitation, cardiovascular exercise, and general fitness programs. Findings indicate that generative AI-generated training programs generally adhere to established exercise guidelines but often lack specificity, progression, and adaptability to real-time physiological feedback. AI-generated recommendations were found to emphasize safety and broad applicability, making them useful for general fitness guidance but less effective for high-performance training. GPT-4 demonstrated superior performance in generating structured resistance training programs compared to older AI models, yet limitations in individualization and contextual adaptation persisted. A critical appraisal using the METRICS checklist revealed inconsistencies in study quality, particularly regarding prompt specificity, model transparency, and evaluation frameworks. While generative AI holds promise for democratizing access to structured exercise prescriptions, its role remains complementary rather than substitutive to expert guidance. Future research should prioritize real-time adaptability, integration with physiological monitoring, and improved AI-human collaboration to enhance the precision and effectiveness of AI-driven exercise recommendations.

Keywords: generative artificial intelligence; large language models; exercise prescription; personalized recommendations; adaptive exercise programs



Academic Editor: Roger Narayan

Received: 23 February 2025

Revised: 13 March 2025

Accepted: 21 March 2025

Published: 22 March 2025

Citation: Puce, L.; Bragazzi, N.L.; Currà, A.; Trompetto, C. Harnessing Generative Artificial Intelligence for Exercise and Training Prescription: Applications and Implications in Sports and Physical Activity—A Systematic Literature Review. *Appl. Sci.* **2025**, *15*, 3497. <https://doi.org/10.3390/app15073497>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Regular physical activity provides significant health benefits [1–3], and exercise and training prescriptions serve as fundamental components of health promotion and athletic performance optimization [4,5]. They are inherently tailored to the diverse needs, goals, and physiological conditions of a wide range of populations, including athletes, either able-bodied or with disabilities, individuals managing chronic conditions, the elderly, and those seeking general well-being [6]. As such, correctly prescribing exercise and devising proper training plans demand expertise and personalized strategies to address the unique characteristics and objectives of each individual [7,8]. When well-designed and evidence-based, exercise and training prescriptions can enhance psychophysical health, optimize biomechanical, psychological, and physiological adaptations to improve performance, and reduce the risk of overuse injuries and adverse health effects [9]. Traditionally, they have relied on the expertise of sports scientists, physiotherapists, and exercise physiologists, using subjective assessments and empirical guidelines. However, the scalability and precision of these traditional methods face limitations, particularly in addressing diverse populations [10,11] or incorporating rapidly evolving scientific evidence into practice [12].

In this context, the integration of artificial intelligence (AI) into healthcare and wellness represents a groundbreaking innovation with far-reaching implications across various domains [13–16], including sports and physical activity [17–20]. AI has demonstrated significant potential to address the limitations of traditional exercise and training prescriptions, offering solutions that are not only scalable but also capable of delivering highly personalized and data-driven recommendations [21,22].

Generative AI, in particular, represents a paradigm shift [23–25]. Encompassing large language models (LLMs) such as ChatGPT, Google Gemini (previously known as Google Bard), and Microsoft Copilot, generative AI has demonstrated exceptional capabilities in conversational interfaces, decision support systems, and personalized recommendations [26–29]. These systems leverage extensive datasets to produce context-aware, adaptive, and human-like responses, making them increasingly valuable in areas requiring individualized guidance, such as exercise and training prescriptions.

By combining advanced natural language processing (NLP) and machine learning (ML) techniques [30,31], generative AI can dynamically generate exercise and training prescriptions that integrate user-specific inputs—such as age, fitness level and training status, health history, and even genetic predispositions—with evidence-based knowledge and sport-specific goals, tailoring recommendations to various timeframes, whether for a single training session, a structured program spanning weeks or months, or a long-term athletic development plan extending over several years.

This would enable the creation of optimized plans customized to various objectives, including general health improvement, rehabilitation, and peak athletic performance [32–35]. Moreover, generative AI has the potential to address critical gaps in accessibility, making professional-level guidance available to individuals who may lack access to specialized expertise or resources due to financial, geographical, or logistical barriers [36–38].

To date, no systematic appraisal exists of the potential applications of generative AI in exercise and training prescription. In light of these considerations, this study aims to critically evaluate the role of generative AI in transforming exercise and training prescription practices. It will explore the strengths, opportunities, and limitations of these technologies, with a particular focus on their application in advanced and high-performance settings. By addressing the challenges and identifying pathways for integration, this analysis seeks to provide a comprehensive understanding of how generative AI can redefine exercise prescriptions for health promotion and training strategies for athletic performance optimization.

2. Methods

2.1. Study Protocol and Study Finding Reporting

The study was conducted following an a priori-specified protocol to ensure transparency and methodological rigor. All aspects of the study, including the objectives, inclusion and exclusion criteria, data collection process, and analysis plan, were defined in advance and documented to prevent deviations that could introduce bias. The findings of the study were reported in accordance with the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses” (PRISMA) checklist [39], which provides a standardized framework to enhance the quality, clarity, and completeness of systematic reviews and meta-analyses. The PRISMA checklist incorporates critical elements such as the study selection process, data extraction methods, risk of bias assessment, and synthesis of results. By adhering to the PRISMA guidelines, the present review ensured that all relevant information was systematically presented, allowing for transparency and reproducibility. This structured approach also facilitates critical appraisal and comparison with other studies and existing reviews in the field, enhancing the overall reliability and credibility of the reported findings.

2.2. Search Strategy

As previously mentioned, to investigate the role of generative AI in exercise prescription, this study adopted a systematic literature review. A comprehensive search strategy was employed to identify relevant studies and reports, utilizing an electronic metadatabase (UnoPerTutto, hosted by Genoa University, Genoa, Italy) [40]. UnoPerTutto integrates multiple scholarly electronic databases, including PubMed/MEDLINE, Scopus, and Web of Science.

Keywords included “generative artificial intelligence”, “large language models”, “chatbots”, “virtual assistants”, “digital assistants”, “physical activity”, “sport”, “exercise prescription”, “training prescription”, “training plan”, “training program”, and related terms, properly connected by means of Boolean operators (Table 1). The inclusion criteria encompassed studies published in peer-reviewed journals and gray literature exploring generative AI systems in the context of sports and physical activity.

2.3. Data Extraction and Data Synthesis

Data extraction focused on the capabilities of generative AI systems in generating meaningful exercise recommendations, the methodologies used to validate their outputs, and their practical applications across different populations. Additionally, the present review analyzed the underlying architectures of AI models, such as transformer-based systems, their ability to integrate domain-specific knowledge, and the type of prompt used (i.e., the input text or query provided to a generative AI system, guiding its response by specifying the context and desired output format). Finally, the ethical, practical, and technical challenges of using generative AI for exercise prescription were noted and critically appraised.

2.4. Critical Appraisal

The critical appraisal of the studies included in this systematic review was conducted using the METRICS checklist [41]. This 9-item evaluation tool, designed to standardize the evaluation of studies involving generative AI models in healthcare education and practice, provided a structured framework for assessing the methodological rigor and reporting quality of each study. The METRICS criteria include key aspects such as the model used, evaluation approach, timing of testing, data transparency, range of tested topics, randomization methods, individual factors in query selection, count of executed queries,

and specificity of prompts and language. Studies meeting 1–5 criteria were classified as low quality, those meeting 6–7 criteria as medium quality, and those meeting 8–9 criteria as high quality. By applying the METRICS checklist, we ensured a systematic and objective appraisal of the included studies, facilitating the identification of strengths and limitations in their design and reporting.

Table 1. An overview of the search strategy adopted in the present systematic review of the literature. Abbreviations: * (wildcard character).

Search Strategy Items	Details
Search string	("generative artificial intelligence" OR "large language model*" OR chatbot* OR "conversational agent" OR "digital assistant" OR "virtual assistant" OR "google bard" OR "google gemini" OR "microsoft copilot" OR chatgpt* OR "generative pre-trained transformer*") AND (sport* OR exercis* OR "physical activity" OR athlet* OR training) AND (prescribing OR prescription* OR recommend*)
Databases	UnoPerTutto hosted by Genoa University, Genoa, Italy
Inclusion Criteria	• Target Populations (P): Studies involving athletes, recreational exercisers, patients with specific medical conditions, or general populations engaging in physical activity • Exposure/Intervention (E/I): Studies or reports explicitly discussing the use of generative AI systems, such as LLMs, chatbots, or virtual assistants (e.g., ChatGPT, Google Bard, Microsoft Copilot), in the context of exercise, physical activity, or sports, i.e., publications that explore generative AI's role in recommending, designing, or tailoring exercise or physical activity programs for individuals or populations • Comparison/Comparator (C): Traditional (human-designed) exercise prescription methods, authoritative guidelines for exercise prescription, and expert opinion • Outcome(s) (O): Effectiveness of generative AI in recommending, designing, or tailoring exercise programs; user adherence and satisfaction with AI-driven exercise guidance; impact on performance, fitness levels, health outcomes, and rehabilitation • Publication Type (S): Peer-reviewed articles, gray literature, and reports from authoritative sports organizations and websites
Exclusion criteria	• Population (P): Studies that do not involve individuals engaged in sports, exercise, or physical activity • Exposure/Interventions (E/I): Studies that do not involve generative AI or focus solely on traditional exercise prescription methods without the integration of AI-based technologies (non-AI interventions); studies relying on pre-generative AI systems or non-transformer-based models that do not meet current technological standards (outdated AI technologies); studies discussing generative AI in contexts unrelated to sports, exercise, or physical activity (e.g., education, entertainment, or general healthcare without an exercise component) (unrelated domains) • Comparison/Comparator (C): Studies that do not compare AI versus non-AI exercise interventions/prescriptions, authoritative guidelines for exercise prescription, or expert opinion • Outcome(s) (O): Articles providing only conceptual or theoretical overviews of generative AI without practical applications or empirical data in exercise prescription (theoretical discussions) • Publication Type (S): Reviews, including systematic reviews, letters to the editor, editorials, comments/commentaries without sufficient data
Language filter	None (any language)
Time filter	None (from inception)

3. Results

Systematic Literature Review Findings

A total of 2891 records were initially identified from database searches. Before the screening, 1619 records were removed as duplicates. No records were excluded based on automation tools or other reasons. Following duplicate removal, 1272 records remained for screening. After screening, 1211 records were excluded, leaving 61 reports for retrieval. All 61 reports were successfully retrieved, with none being unavailable. Subsequently, 61 reports were assessed for eligibility. Ultimately, 10 studies [42–51] were included in the systematic review. Further details are pictorially shown in Figure 1.

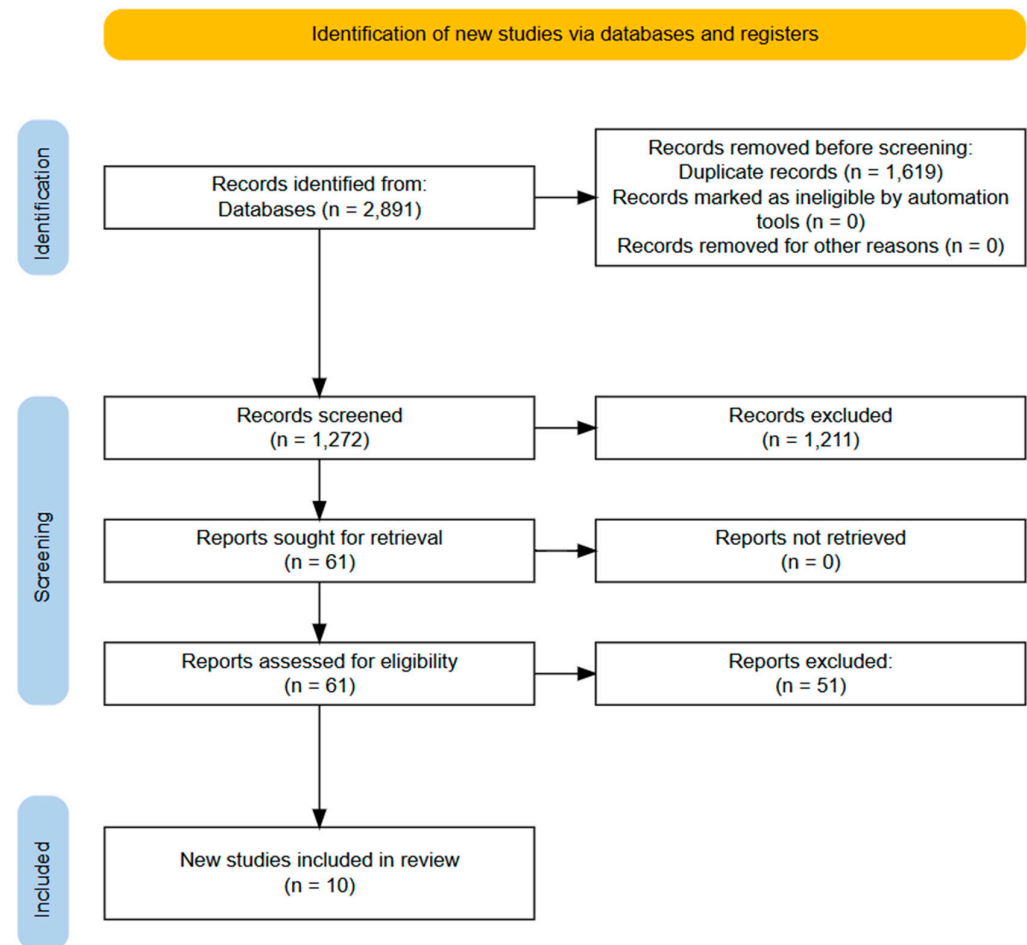


Figure 1. Pictorial flowchart of the present systematic review and meta-analysis.

The studies represented a diverse geographic distribution across North America, Europe, Asia, and Africa. Among them, one study involved a multi-national collaboration, with authors from 20 countries [43], while others were conducted mainly in Germany [44,46], Turkey [45], China [49,51], Thailand [47], the Philippines [42], and the USA [50]. Finally, a study was multi-country [48].

The studies employed a range of methodologies, including qualitative/quasi-qualitative assessments and expert panel reviews [43,48,51], quantitative assessments [45], mixed-methods approaches [44,46,49,51], a quasi-experimental design [42], and a randomized controlled trial (RCT) [47], reflecting the evolving nature of research on AI-generated exercise prescription.

The studies primarily examined the capabilities of different AI models in generating exercise programs [43,47,49–51] or training plans [29,44–46,48] and their performance relative to human experts.

ChatGPT was the most frequently evaluated AI system, with various versions included across eight studies. GPT-4 was assessed in several studies [43,46–49,51], while ChatGPT-3.5 was compared in others [42,45,48,50]. One study [46] included Google Gemini, finding that GPT-4 outperformed it in training plan quality. One study [44] used ChatGPT-3. Finally, another study [49] compared ChatGPT to the Intelligent Health Promotion Systems (IHPS), demonstrating that ChatGPT excelled in accuracy and comprehensiveness, although IHPS had greater consistency in applicability.

The study populations varied significantly, ranging from expert evaluations without real participants [43–46] to a few empirical studies [42,47,49]. Two of these involved small samples, with five [49] and nine [47] participants, respectively. Larger participant groups were included in [42], which examined 87 individuals. Demographic considerations were generally limited, with the exception of a single study [50] that analyzed exercise recommendations across 26 diverse populations. These included healthy adults, older individuals, pregnant individuals, and those with chronic conditions such as cardiovascular disease, diabetes, cancer, and HIV.

The AI-generated exercise prescriptions were assessed using various frameworks, including adherence to the FITT standardized training principles such as periodization, training volume, and intensity [43,46,48,50,51], or the ACSM guidelines [50,51]. Additionally, evaluations considered factors like accuracy and adequacy [45,49,50]. Notably, the criteria used for assessment were generally limited, with only one study [44] analyzing a comprehensive set of 22 criteria. Lastly, only a few studies [42,47] employed objective measures to validate the reliability of AI-generated exercise plans.

The AI-generated programs covered various exercise domains, including core exercises [45], strength training [46,48] running plans [44], calisthenics [42], weight loss exercises [47], and rehabilitation [43,49–51].

Overall, the findings suggest that AI-generated exercise programs align with general exercise science principles but often lack specificity, progression, and adaptability. Studies that assessed the quality of AI-generated programs against expert-designed interventions consistently reported that AI output required expert validation and modifications to optimize effectiveness. While ChatGPT-created plans adhered to foundational training principles, they were frequently generic and did not incorporate real-time feedback or individualized adjustments. AI-generated running and strength training plans improved when additional user input was provided, but they still fell short of human expert-designed programs [44,46,48]. Furthermore, while AI-generated programs generally adhered to recognized exercise guidelines, some studies identified limitations in specificity, including an overemphasis on safety and moderate intensity at the expense of progression and individualization [43,50].

In studies that examined physiological outcomes, AI-generated programs demonstrated positive effects in certain domains. Improvements were observed in BMI reduction, heart rate response, and muscular endurance [42,47]. However, in direct comparisons with human-designed interventions, AI-generated exercise prescriptions were less effective in improving cardiovascular endurance and lower-body muscular strength, particularly in males [42]. Additionally, some AI-generated recommendations displayed biases, including an excessive emphasis on medical clearance requirements, which disproportionately affected older adults and individuals with disabilities [36].

When comparing different AI models, GPT-4 demonstrated superior performance relative to GPT-3.5 in generating tailored and structured resistance training programs,

incorporating more advanced training concepts such as block periodization and active recovery [48]. The comparison between Google Gemini and GPT-4 showed that GPT-4 produced higher-quality training plans regardless of input specificity [46]. Nonetheless, reproducibility varied, and expert oversight was deemed essential to ensure the quality and applicability of AI-generated exercise recommendations.

Further details are presented in Table 2.

In terms of critical appraisal (Table 3), the studies analyzed exhibit varying degrees of methodological rigor and comprehensiveness, particularly concerning model specification, evaluation approach, timing of testing, transparency of data sources, and overall quality. A few studies, such as those by Düking et al. [44], Havers et al. [46], and Zaleski et al. [50], demonstrate high methodological robustness, incorporating systematic topic selection, strong interrater reliability, and high prompt specificity. These studies also exhibit transparency in their data sources, ensuring replicability and credibility. In contrast, studies such as Erol and Arıkan [45], Philuek et al. [47], Xu et al. [49], and Zhu et al. [51] display notable methodological weaknesses, particularly in terms of transparency, topic selection, and reliability. The lack of systematic topic selection and insufficient prompt specificity in these studies contribute to their lower overall quality, limiting their applicability and generalizability. Other studies, such as those by Dergaa et al. [43] and Washif et al. [48], occupy an intermediate position, meeting several methodological criteria but lacking in aspects such as topic selection and interrater reliability.

A key differentiating factor among the studies is the range of tested topics and whether topic selection followed a systematic or randomized approach. Studies with a systematic approach tend to yield more reliable and reproducible findings, as demonstrated by Düking et al. [44] and Havers et al. [46]. Conversely, those lacking a structured topic selection process, such as Philuek et al. [47] and Zhu et al. [51], exhibit greater methodological limitations. Prompt specificity emerges as another critical determinant of study quality. While high-quality studies employ precise and well-structured prompts, lower-quality studies, such as those by Xu et al. [49] and Zhu et al. [51], fail to provide adequate specificity, leading to less reliable conclusions. Additionally, transparency in data sources plays a fundamental role in the overall robustness of a study. Those who lack explicit disclosure of their data sources, such as Erol and Arıkan [45] and Masagca [42], suffer from reduced credibility. Overall, studies that incorporate rigorous model specification, transparent methodologies, systematic topic selection, and high interrater reliability tend to produce more reliable findings. Those who neglect these aspects exhibit lower methodological soundness, reducing their contribution to the field.

Table 2. An overview of the studies included in the present systematic review of the literature.

Study	Study Country	Study Design	Sample Size	Evaluation Framework and Systems Compared	Exercise Prescription	Key Metrics	Main Findings
Dergaa et al., 2024 [43]	20 countries	Quasi-qualitative assessment and expert panel review	Five hypothetical patient scenarios (hypertension, osteoarthritis, anxiety, type 2 diabetes, asthma; 2 males, 3 females, age 27–50 years)	GPT-4 vs. 38 specialists in sports medicine, exercise science, and rehabilitation	Standardized prompt asking GPT-4 to create a 30-day exercise program using the FITT principles	Adherence to FITT principles, integration of perceived exertion, safety considerations, and individualization	GPT-4 created safe, general exercise programs, which lacked specificity and progression for individual health conditions Other key limitations include overemphasis on safety and moderate intensity, lack of real-time feedback and monitoring, generic approach due to the single-interaction design
Düking et al., 2024 [44]	Germany, Netherlands, UK	Mixed-methods approach (expert panel review, quasi-qualitative assessments)	A fictional 20-year-old male runner	ChatGPT (version 3.0.1) vs. 10 experienced coaches with seven years of experience	6-week running plans incorporating intervals, long runs, and recovery	22 criteria (18 primary and 4 secondary)	Plans with more input received higher quality ratings However, AI-generated plans lack direct interactions and feedback with users and are not completely evidence-based, requiring expert validation
Erol and Arıkan, 2024 [45]	Turkey	Quantitative assessments	Not applicable	ChatGPT-3.5 vs. nine experienced physiotherapists with 6–11 years of experience	23 knowledge items related to core exercises	Accuracy/adequacy	ChatGPT was generally satisfactory in providing answers related to core exercises Best performance was in general knowledge questions, while it struggled with individualized programming and specific recommendations
Havers et al., 2025 [46]	Germany	Mixed-methods approach (expert panel review, quasi-qualitative assessment)	A fictitious person	Google Gemini 1.0 Pro and GPT-4 vs. 12 coaching experts with at least 3 years of experience	8-week muscle hypertrophy-related resistance training plans	Key training aspects covering exercise selection, training intensity, weekly frequency, repetition range, and recovery principle	More detailed input improved LLM-generated plans, but coaching experts still rated them below optimal levels GPT-4 outperformed Google Gemini in training plan quality, regardless of input detail Reproducibility varied

Table 2. Cont.

Study	Study Country	Study Design	Sample Size	Evaluation Framework and Systems Compared	Exercise Prescription	Key Metrics	Main Findings
Masagca, 2025 [42]	Philippines	Quasi-experimental; one-group pre-test-post-test for within-group comparison and two-group pre-test-post-test for between-group comparison	87 untrained collegiate students (44 females, 43 males); 43 in the AI-generated calisthenics training program (AIGCTP) group, 44 in the human-made calisthenics training program (HMCTP) group	ChatGPT-3.5	Prompt based on FITT principles—a 5-week calisthenics training program, including flexibility, cardiovascular endurance, and muscular endurance components	Flexibility (sit and reach test), cardiovascular endurance (3-min step test), muscular endurance (wall sit, plank, and push-up tests)	AIGCTP significantly improved lower extremity flexibility and upper extremity muscular endurance in males but had limited impact on females HMCTP showed improvements in cardiovascular endurance, lower limb flexibility, and muscular endurance of upper and lower extremities in males The traditional program outperformed AI-generated training in cardiovascular endurance and some male-specific metrics
Philuek et al., 2025 [47]	Thailand	Randomized controlled trial (intervention study)	9 participants aged 19 years (ChatGPT-generated exercise program: 6; Control: 3)	ChatGPT-4.0	Exercise for weight reduction—8-week program, 3 sessions per week (a 5–10-min warm-up, 45–60 min of physical fitness exercises—aerobic, resistance training, flexibility –, and a 5–10-min cool-down)	BMI, percent of fat, level of visceral fat, basal metabolic rate, percent of skeletal muscle, percent of subcutaneous fat, heart rate after standing and knee lifting for 3 min, hand grip strength, sit and stand in 30 s, flexibility, and lung capacity	The ChatGPT group showed significant improvements in BMI, heart rate after standing and knee lifting, and sit-and-stand repetitions in 30 s

Table 2. Cont.

Study	Study Country	Study Design	Sample Size	Evaluation Framework and Systems Compared	Exercise Prescription	Key Metrics	Main Findings
Washif et al., 2024 [48]	Malaysia, Czech Republic, Hong Kong, Qatar, Tunisia, New Zealand	Qualitative assessments	A hypothetic male and female individual, aged 20 years, with intermediate and advanced resistance training experience	ChatGPT-3.5 and -4.0 vs. established guidelines (e.g., National Strength and Conditioning Association textbook)	Standardized instructions requesting 12-week resistance training programs for specific experience levels	Periodization, exercise selection, training volume, load intensity, tempo, rest intervals, and progression	GPT 4.0 generated more comprehensive and tailored programs than GPT 3.5, considering advanced training principles like block periodization and active recovery However, programs required expert modification to align with best practices Key limitations included lack of real-time adaptability, emerging methodologies (e.g., blood flow restriction), and sex-specific guidance
Xu et al., 2024 [49]	China	Mixed-methods approach with patient data collected via questionnaires and hardware tools	5 hypertensive patients (3 females, 2 males) with comorbidities, aged 69–79 years, with conditions such as diabetes, COPD, chronic nephritis, Parkinson’s disease, and gouty arthritis	ChatGPT-4.0 and Intelligent Health Promotion Systems (IHPS) vs. 24 multidisciplinary experts from over ten different professional fields, with more than 10 years of experience	Exercise prescription for hypertensive patients based on expected health benefits, FITT principles, and safety	Accuracy, comprehensiveness, applicability, and evaluation based on the Transtheoretical Model	ChatGPT outperformed IHPS in accuracy and comprehensiveness, but IHPS had better applicability consistency ChatGPT did not take into account cultural preferences and delivered standardized, repetitive prescriptions Gaps in medication management, adaptability, and personalization

Table 2. Cont.

Study	Study Country	Study Design	Sample Size	Evaluation Framework and Systems Compared	Exercise Prescription	Key Metrics	Main Findings
Zaleski et al., 2024 [50]	USA	Mixed-methods approach (conceptual content analysis and thematic mapping)	26 populations across the lifespan including healthy adults, older adults, children and adolescents, pregnant individuals, and those with chronic diseases such as CVD, diabetes, cancer, and HIV	ChatGPT-3.5 vs. ACSM guidelines	Exercise recommendations for diverse populations	Accuracy, comprehensiveness/depth (adherence to the FITT principles, and alignment with ACSM guidelines), and readability	Moderate comprehensiveness and high accuracy, with low readability and gaps in exercise frequency, intensity, time, and volume guidance, misinformation (e.g., medical clearance overemphasis for preparticipation screening), inconsistencies in the terminology used for exercise professionals, liability concerns leading to bias toward safety, and discrimination against age-based and disabled populations
Zhu et al., 2024 [51]	China and USA	Qualitative assessments, with case studies—two cases (a case from the ACSM Guidelines and a fictional case)	Patients undergoing post-stent cardiac rehabilitation (a 60-year-old woman) and with Parkinson’s disease	ChatGPT-4.0 vs. ACSM guidelines	Medical clearance for a professionally led walking program and an aerobic and strength training program	Adherence to the FITT-VP principles, alignment with ACSM guidelines	ChatGPT aligned with ACSM guidelines and provided additional context (e.g., balance, safety, and motivation)

Table 3. Critical appraisal of the studies included in the present systematic review.

Study	Model Specification	Evaluation Approach	Timing of Testing	Transparency of Data Source	Range of Tested Topics	Topic Selection (Randomized/Systematic)	Interrater Reliability/Reliability	Number of Queries	Prompt Specificity	Overall Quality
Dergaa et al., 2024 [43]	✓	✓	✓	✓	✓	✗	✗	✓	✓	Medium
Düking et al., 2024 [44]	✓	✓	✓	✓	✓	✓	✓	✓	✓	High
Erol and Arikan, 2024 [45]	✓	✓	✗	✗	✗	✓	✗	✓	✗	Low
Havers et al., 2025 [46]	✓	✓	✓	✓	✓	✓	✓	✓	✓	High
Masagca, 2025 [42]	✓	✓	✗	✓	✗	✗	✓	✗	✓	Medium
Philuek et al., 2025 [47]	✓	✓	✗	✓	✗	✗	✓	✗	✗	Low
Washif et al., 2024 [48]	✓	✓	✓	✓	✓	✗	✗	✓	✓	Medium
Xu et al., 2024 [49]	✓	✓	✗	✗	✗	✗	✓	✗	✓	Low
Zaleski et al., 2024 [50]	✓	✓	✓	✓	✓	✓	✓	✓	✗	High
Zhu et al., 2024 [51]	✓	✓	✗	✓	✗	✗	✗	✗	✗	Low

4. Discussion

According to the studies included in the present systematic literature review, AI-generated exercise prescriptions show promise but remain limited by a lack of real-time adaptability, individualization, and direct user feedback. While GPT-4 demonstrated notable improvements over earlier models and competing AI systems, AI-generated programs still require expert modifications to align with best practices. Although AI-driven approaches offer an innovative complement to exercise planning, their effectiveness in real-world settings will depend on future advancements in adaptive capabilities and integration with human expertise.

4.1. The Role of Generative AI in Personalized Exercise Programs for Health and Disease

Dergaa et al. [43], Philuek et al. [47], Xu et al. [49], Zaleski et al. [50], and Zhu et al. [51] collectively provide a nuanced evaluation of AI models' capabilities in exercise prescription across different methodologies, populations, and healthcare frameworks. While each study approached the evaluation from distinct perspectives, common themes emerged regarding AI models' strengths and limitations. One recurring observation across all studies is AI models' strong emphasis on safety and general adherence to established guidelines. Dergaa et al. [43] found that GPT-4 produced safe but generic exercise programs, with insufficient progression and condition-specific tailoring. This aligns with Zaleski et al. [50], who reported that ChatGPT-3.5 frequently prioritized liability concerns over individualized exercise prescriptions, particularly for vulnerable populations such as older adults and those with chronic diseases. Similarly, Xu et al. [49] noted that GPT-4 performed well in accuracy and comprehensiveness when prescribing exercise for hypertensive patients with comorbidities but lagged behind IHPS in applicability and scenario-specific consistency. At the same time, the studies differ in how they assess GPT-4's adaptability to existing guidelines. While Dergaa et al. [43] and Zaleski et al. [50] highlight gaps in specificity and individualization, Zhu et al. [51] provide a more optimistic outlook, noting that GPT-4 generally aligns well with ACSM guidelines for cardiac rehabilitation and Parkinson's

disease management. However, Zhu et al. [51] also emphasize that AI-generated recommendations remain complementary rather than standalone, requiring clinical interpretation and integration by healthcare professionals. This point resonates with Xu et al. [49], who suggest that GPT-4 is most effective when used in conjunction with expert judgment rather than as a replacement for human decision-making.

Philuek et al.'s RCT [47] contributes to this discussion, showing that the ChatGPT group demonstrated significant improvements in BMI, heart rate after standing and knee lifting, and sit-and-stand performance, and suggesting that AI-generated exercise programs can drive meaningful physiological benefits in young adults.

A critical difference across studies lies in how AI models' limitations manifest depending on the evaluation framework. Dergaa et al. [43] and Zaleski et al. [50] underscore AI models' conservative approach, often erring on the side of safety at the cost of personalization and progression. Zaleski et al. [50] also highlight potential biases, particularly regarding age and disability, where ChatGPT-3.5's cautious recommendations may inadvertently reinforce disparities. In contrast, Xu et al. [49] provide a more comparative lens, demonstrating that while GPT-4 is effective in accuracy and comprehensiveness, it is less reliable in practical implementation compared to specialized systems like IHPS. The findings from Philuek et al. [47] provide additional evidence supporting GPT-4's ability to generate structured exercise programs that lead to tangible physiological improvements, although the small sample size limits the generalizability of the findings.

Taken together, these studies indicate that while AI models are valuable tools for generating exercise recommendations based on established guidelines, their outputs require further refinement to enhance specificity, progression, and real-world applicability. The findings emphasize the importance of human oversight in exercise prescription, particularly in addressing individual needs, mitigating biases, and ensuring that AI-driven recommendations are effectively integrated into clinical practice.

4.2. The Role of Generative AI in Personalized Exercise Programs for Athletes

Düking et al. [44], Erol and Arıkan [45], Havers et al. [46], Masagca [42], and Washif et al. [48] collectively explore the capabilities and limitations of ChatGPT and other LLMs in generating training programs across diverse contexts.

Despite differences in methodology, study design, and target populations, a common theme emerges: while AI-generated programs can align with foundational training principles, they often lack the depth, specificity, and adaptability required for personalized and high-performance training. A key insight from Düking et al. [44] is that the granularity of user input significantly influences the quality of ChatGPT-generated training plans. Their study demonstrated that a six-week running program improved in quality when detailed user-specific data were provided. However, even the most refined AI-generated plans had critical gaps, particularly in health screening, training frequency progression, psychological skills training, and skill acquisition. This finding aligns with Erol and Arıkan [45], who assessed ChatGPT's accuracy in answering questions about core exercises. They found that while ChatGPT excelled in general knowledge responses, it struggled with individualized recommendations and contraindications—reinforcing Düking et al. [44]'s conclusion that AI-generated outputs lack sufficient customization and real-time adaptability. Masagca [42] took this inquiry a step further by conducting a quasi-experimental study to compare AI-generated and human-designed calisthenics training programs. While the AI-generated program improved flexibility and upper extremity endurance in male participants, it was notably less effective than the human-made program in enhancing cardiovascular endurance and lower extremity strength. Female participants in the AI-generated group showed no significant improvements at all. These findings suggest that, while AI can assist

in exercise programming, its efficacy varies across populations and fitness components—a limitation also noted by Düking et al. [44], whose study revealed ChatGPT's shortcomings in monitoring contextual factors such as environmental conditions and individual load responses. Beyond individual studies, comparative analyses of different LLMs provide additional context. Washif et al. [48] highlighted GPT-4's superiority over GPT-3.5 in designing resistance training programs, particularly in implementing advanced programming concepts such as block periodization. However, GPT-4 still fell short in accommodating real-time adaptability, sex-specific considerations, and emerging methodologies like blood flow restriction training.

Havers et al. [46] further explored LLM-based resistance training plans, comparing GPT-4 with Google Gemini. Their findings revealed significant variability in output quality, with GPT-4 producing more structured and evidence-based programs, especially when given detailed input. Google Gemini, in contrast, struggled to generate coherent hypertrophy training plans when minimal input was provided, emphasizing the importance of prompt specificity. Both studies highlight a broader issue: while AI can assist in structured programming, reproducibility, and consistency remain challenges, as identical prompts sometimes yield different outputs.

Taken together, these studies paint a nuanced picture of AI's role in exercise programming. While LLMs like GPT-4 can generate structured plans based on established principles, their effectiveness is highly dependent on input granularity. These systems heavily depend on the development of robust methods and precisely engineered prompts, which play a crucial role in enhancing accuracy, ensuring safety, and aligning AI-generated recommendations with real-world physiological and biomechanical principles. The optimal design, structure, and contextual adaptation of prompts—including how variations in wording, specificity, and data integration influence the quality, relevance, and personalization of AI-generated exercise and training prescriptions—are essential for ensuring reliability and reproducibility [52,53].

Additionally, AI-generated programs often lack essential elements such as individualized health screening, real-time adaptability, and nuanced periodization strategies. Comparative studies further underscore that while GPT-4 currently leads in generating structured programs, variability and consistency issues remain prevalent across LLMs.

Moreover, the capability of these systems to effectively integrate multidisciplinary data, adapt to users' dynamic needs, and complement human expertise requires further investigation [54–56]. Furthermore, while AI excels in creating baseline programs and prioritizing general safety, it often lacks the nuanced adaptability required to meet the complex and evolving needs of advanced athletes. For this population, exercise and training prescriptions must account for highly individualized and progressive adaptations in training load, intensity, and periodization [57–60], as well as the integration of multidisciplinary factors such as nutrition, recovery strategies, mental health, and injury prevention [61,62]. Current AI systems are also limited in their ability to provide real-time feedback or to incorporate contextual factors, such as environmental conditions or acute physiological responses, which are critical for optimizing outcomes in elite or specialized training contexts [63,64]. For instance, Rocha-Silva et al. [64] evaluated the accuracy of information provided by ChatGPT (versions 3.5 and 4o) regarding the role of lactate in fatigue and muscle pain during physical exercise. Both versions incorrectly attributed fatigue to glycogen depletion and lactic acid accumulation, while linking pain to inflammation and microtrauma. Further interactions with ChatGPT, including user feedback and refined prompts, led to more accurate responses that debunked the misconception of lactate as the primary cause of fatigue.

These limitations underscore the need for human oversight to ensure that AI-generated programs align with the specific demands of advanced users and high-performance scenarios [65], suggesting that AI-generated exercise prescriptions should be viewed as complementary tools rather than substitutes for expert-designed training regimens.

4.3. Future Directions

The findings of the studies included in this systematic review are limited by the small sample sizes in many studies—often with few or no real participants—which restricts the generalizability of the results.

To advance the role of generative AI in exercise prescription, future research must explore larger, more diverse populations, incorporating real-world data to validate AI-generated exercise recommendations. Further, it should prioritize several key areas. First, it should explore the integration of real-time adaptability mechanisms, which is essential for improving the specificity and progression of AI-generated recommendations. This could involve the use of cutting-edge techniques such as reinforcement learning or hybrid models that combine generative AI with sensor-driven data from wearable devices, enabling continuous monitoring and adaptation based on an individual's performance and physiological responses [54–56,63,64]. By leveraging real-time biometrics, such as heart rate variability, muscle fatigue markers, and motion analysis, AI models can dynamically adjust exercise intensity, duration, and recovery periods to optimize training effectiveness while minimizing injury risks. Additionally, these adaptive systems can personalize recommendations based on an individual's unique physiological traits, ensuring that exercise programs are not only inclusive but also tailored to the specific needs of diverse populations, including female athletes and individuals with disabilities.

This can be achieved by incorporating advancements in exercise science, including novel training methodologies and population-specific adaptations. This necessitates collaboration between AI developers and domain experts to ensure that emerging evidence is seamlessly integrated into the AI's knowledge base [66]. Additionally, efforts to minimize biases and enhance inclusivity must be central to the development process. This includes addressing age, gender, and disability-related disparities in AI recommendations and ensuring that these systems are equitable and accessible to all [67]. Only a few studies explored these aspects, including sex-specific differences in AI-generated exercise prescriptions. For instance, the study by Masagca [42] revealed that AI-generated calisthenics training programs had a differential impact on male and female participants, with significantly better improvements observed in males for flexibility and muscular endurance compared to females. This suggests that current generative AI systems may not adequately incorporate sex-specific physiological and biomechanical differences when designing exercise regimens. Several factors could contribute to this limitation. AI models trained on general fitness datasets may underrepresent sex-based variations in metabolic responses, muscle composition, hormonal influences on recovery, and biomechanical differences affecting exercise execution. The absence of such considerations could lead to suboptimal or even counterproductive recommendations for female athletes and recreational exercisers.

Moreover, the ethical and regulatory landscape surrounding AI in exercise prescription and healthcare requires significant attention [67]. Establishing robust frameworks for data privacy, accountability, and liability is critical to building trust among users and stakeholders. Given that AI-driven exercise recommendations rely on vast amounts of personal health data, ensuring compliance with global data protection laws such as GDPR and HIPAA is paramount. This includes implementing stringent encryption protocols, secure data-sharing agreements, and transparent consent mechanisms to protect user privacy. Additionally, clear guidelines must be developed to define the scope of AI's responsibility

versus human oversight in exercise prescription, ensuring that healthcare providers and fitness professionals retain decision-making authority while benefiting from AI-driven insights. Liability concerns also need to be addressed, particularly in cases where AI-generated recommendations lead to adverse effects or injuries. Regulatory bodies should work alongside AI developers and medical professionals to establish standardized risk assessment protocols and ethical guidelines, ensuring that AI-enhanced exercise plans are not only effective but also safe, equitable, and unbiased. Furthermore, ongoing monitoring and validation of AI models should be mandated to prevent algorithmic drift and ensure that recommendations remain aligned with the latest scientific evidence. Public transparency initiatives, such as AI explainability features, can further enhance trust by allowing users to understand the rationale behind AI-generated exercise plans, fostering greater adoption and confidence in these technologies.

Furthermore, interdisciplinary efforts should focus on designing AI systems that complement, rather than replace, human expertise, and longer-term experimental studies are needed to assess whether AI-driven programs can evolve to match the efficacy of human-designed training over extended periods. In this regard, rigorous validation studies comparing AI-generated recommendations with human expertise and patient outcomes are necessary to substantiate the clinical and practical utility of these tools. In addition to these comparative studies, generative AI itself could be leveraged as a validation tool by assessing proposed exercise plans against the latest scientific criteria and evidence-based guidelines. By analyzing vast repositories of sports science literature, clinical exercise recommendations, and biomechanical data, generative AI can provide real-time evaluations of exercise regimens, highlighting potential discrepancies, missing elements, or biases in the plan. This approach would not only enhance the reliability of AI-generated prescriptions but also support fitness professionals and clinicians in refining exercise interventions to align with the most current best practices.

Thus, generative AI has the potential to function as a collaborative tool, supporting sports scientists, physiotherapists, and exercise physiologists by automating routine tasks, synthesizing complex datasets, and enhancing decision-making processes. However, this requires a clear delineation of roles, ensuring that AI systems are deployed as part of an integrated, human-centered approach to exercise prescription.

5. Conclusions

Generative AI represents a promising frontier in the fields of exercise prescription, sports science, and public health, offering unprecedented opportunities to enhance the precision, scalability, and accessibility of individualized exercise guidance.

The present systematic literature review underscores the potential of generative AI systems like GPT-4 to adhere to established frameworks and principles while identifying key limitations that currently hinder their broader applicability. The lack of specificity, progression, and real-time adaptability, along with biases in addressing diverse populations, highlights the need for ongoing refinement and multidisciplinary collaboration.

While generative AI tools are unlikely to replace human expertise, they hold immense potential as complementary systems capable of democratizing access to high-quality exercise guidance. By addressing the technical, ethical, and practical challenges identified in this study, future developments could pave the way for AI-driven solutions that not only meet the needs of elite athletes but also promote health and well-being across diverse populations. In doing so, generative AI has the potential to revolutionize exercise prescription, bridging the gap between evidence-based guidelines and personalized care in an increasingly digital healthcare landscape.

Author Contributions: Conceptualization, L.P. and N.L.B.; methodology, N.L.B.; software, N.L.B.; validation, L.P., N.L.B., A.C. and C.T.; formal analysis, N.L.B.; investigation, L.P. and N.L.B.; resources, N.L.B.; data curation, N.L.B.; writing—original draft preparation, L.P. and N.L.B.; writing—review and editing, L.P., N.L.B., A.C. and C.T.; visualization, N.L.B.; supervision, C.T.; project administration, A.C.; funding acquisition, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Warburton, D.E.R.; Nicol, C.W.; Bredin, S.S.D. Health Benefits of Physical Activity: The Evidence. *Can. Med. Assoc. J.* **2006**, *174*, 801–809. [\[CrossRef\]](#)
- Reiner, M.; Niermann, C.; Jekauc, D.; Woll, A. Long-term health benefits of physical activity—a systematic review of longitudinal studies. *BMC Public Health* **2013**, *13*, 813. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
- Ammar, A.; Trabelsi, K.; Hermassi, S.; Kolahi, A.-A.; Mansournia, M.; Jahrami, H.; Boukhris, O.; Boujelbane, M.; Glenn, J.; Clark, C.; et al. Global Disease Burden Attributed to Low Physical Activity in 204 Countries and Territories from 1990 to 2019: Insights from the Global Burden of Disease 2019 Study. *Biol. Sport* **2023**, *40*, 835–855. [\[CrossRef\]](#) [\[PubMed\]](#)
- Phillips, E.M.; Kennedy, M.A. The exercise prescription: A tool to improve physical activity. *PM&R* **2012**, *4*, 818–825. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wackerhage, H.; Schoenfeld, B.J. Personalized, Evidence-Informed Training Plans and Exercise Prescriptions for Performance, Fitness and Health. *Sports Med.* **2021**, *51*, 1805–1813. [\[CrossRef\]](#) [\[PubMed\]](#)
- Buford, T.W.; Roberts, M.D.; Church, T.S. Toward exercise as personalized medicine. *Sports Med.* **2013**, *43*, 157–165. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
- Lehtonen, E.; Gagnon, D.; Eklund, D.; Kaseva, K.; Peltonen, J.E. Hierarchical framework to improve individualised exercise prescription in adults: A critical review. *BMJ Open Sport Exerc. Med.* **2022**, *8*, e001339. [\[CrossRef\]](#) [\[PubMed\]](#) [\[PubMed Central\]](#)
- Lucini, D.; Pagani, M. Exercise Prescription to Foster Health and Well-Being: A Behavioral Approach to Transform Barriers into Opportunities. *Int. J. Environ. Res. Public Health* **2021**, *18*, 968. [\[CrossRef\]](#)
- Rooney, D.; Gilmartin, E.; Heron, N. Prescribing Exercise and Physical Activity to Treat and Manage Health Conditions. *Ullst. Med. J.* **2023**, *92*, 9–15.
- Milani, J.G.P.O.; Milani, M.; Verboven, K.; Cipriano, G.; Hansen, D. Exercise Intensity Prescription in Cardiovascular Rehabilitation: Bridging the Gap between Best Evidence and Clinical Practice. *Front. Cardiovasc. Med.* **2024**, *11*, 1380639. [\[CrossRef\]](#)
- Almarcha, M.; Sturmberg, J.; Balagué, N. Personalizing the Guidelines of Exercise Prescription for Health: Guiding Users from Dependency to Self-Efficacy. *Apunt. Sports Med.* **2024**, *59*, 100449. [\[CrossRef\]](#)
- Zenko, Z.; Ekkekakis, P. Knowledge of Exercise Prescription Guidelines Among Certified Exercise Professionals. *J. Strength Cond. Res.* **2015**, *29*, 1422–1432. [\[CrossRef\]](#) [\[PubMed\]](#)
- Vyas, S.; Gupta, S.; Shukla, V.K. Towards Edge AI and Varied Approaches of Digital Wellness in Healthcare Administration: A Study. In Proceedings of the 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 9–10 March 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 186–190.
- Al Kuwaiti, A.; Nazer, K.; Al-Reedy, A.; Al-Shehri, S.; Al-Muhanna, A.; Subbarayalu, A.V.; Al Muhanna, D.; Al-Muhanna, F.A. A Review of the Role of Artificial Intelligence in Healthcare. *J. Pers. Med.* **2023**, *13*, 951. [\[CrossRef\]](#)
- Iqbal, J.; Jaimes, D.C.C.; Makineni, P.; Subramani, S.; Hemaida, S.; Thugu, T.R.; Butt, A.N.; Sikto, J.T.; Kaur, P.; Lak, M.A.; et al. Reimagining healthcare: Unleashing the power of artificial intelligence in medicine. *Cureus* **2023**, *15*, e44658.
- Sezgin, E. Artificial intelligence in healthcare: Complementing, not replacing, doctors and healthcare providers. *Digit. Health* **2023**, *9*, 20552076231186520. [\[PubMed\]](#)
- Reis, F.J.J.; Alaiti, R.K.; Vallio, C.S.; Hespanhol, L. Artificial Intelligence and Machine Learning Approaches in Sports: Concepts, Applications, Challenges, and Future Perspectives. *Braz. J. Phys. Ther.* **2024**, *28*, 101083. [\[CrossRef\]](#)
- Mateus, N.; Abade, E.; Coutinho, D.; Gómez, M.-Á.; Peñas, C.L.; Sampaio, J. Empowering the Sports Scientist with Artificial Intelligence in Training, Performance, and Health Management. *Sensors* **2024**, *25*, 139. [\[CrossRef\]](#)

19. Krstić, D.; Vučković, T.; Dakić, D.; Ristić, S.; Stefanović, D. The application and impact of artificial intelligence on sports performance improvement: A systematic literature review. In Proceedings of the 2023 4th International Conference on Communications, Information, Electronic and Energy Systems (CIEES), Plovdiv, Bulgaria, 23–25 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–8.
20. Naughton, M.; Salmon, P.M.; Compton, H.R.; McLean, S. Challenges and Opportunities of Artificial Intelligence Implementation within Sports Science and Sports Medicine Teams. *Front. Sports Act. Living* **2024**, *6*, 1332427. [CrossRef]
21. Chen, H.-K.; Chen, F.-H.; Lin, S.-F. An AI-Based Exercise Prescription Recommendation System. *Appl. Sci.* **2021**, *11*, 2661. [CrossRef]
22. Balpande, M.; Sharma, J.; Nair, A.; Khandelwal, M.; Dhanray, S. AI Based Gym Trainer and Diet Recommendation System. In Proceedings of the 2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON), Mysore, India, 5–7 August 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–7.
23. Reddy, S. Generative AI in Healthcare: An Implementation Science Informed Translational Path on Application, Integration and Governance. *Implement. Sci.* **2024**, *19*, 27. [CrossRef]
24. Fayed, A.M.; Mansur, N.S.B.; de Carvalho, K.A.; Behrens, A.; D’hooghe, P.; Netto, C.d.C. Artificial Intelligence and ChatGPT in Orthopaedics and Sports Medicine. *J. Exp. Orthop.* **2023**, *10*, 74. [CrossRef] [PubMed]
25. Oliver, A.; Guiller, J. Generative AI in Sport and Exercise Psychology: Exploring Opportunities and Overcoming Challenges. *Sport Exerc. Psychol. Rev.* **2025**, *19*, 36–45. [CrossRef]
26. Chang, Y.; Wang, X.; Wang, J.; Wu, Y.; Yang, L.; Zhu, K.; Chen, H.; Yi, X.; Wang, C.; Wang, Y.; et al. A Survey on Evaluation of Large Language Models. *ACM Trans. Intell. Syst. Technol.* **2024**, *15*, 1–45. [CrossRef]
27. Naveed, H.; Khan, A.U.; Qiu, S.; Saqib, M.; Anwar, S.; Usman, M.; Akhtar, N.; Barnes, N.; Mian, A. A Comprehensive Overview of Large Language Models. *arXiv* **2024**, arXiv:2307.06435.
28. Solomon, T.; Laye, M. Examining the Sports Nutrition Knowledge of Large Language Model. (LLM) Artificial Intelligence (AI) Chatbots. 2024. Available online: <https://osf.io/zckya/resources> (accessed on 13 March 2025).
29. Puce, L.; Ceylan, H.I.; Trompetto, C.; Cotellessa, F.; Schenone, C.; Marinelli, L.; Zmijewski, P.; Bragazzi, N.; Mori, L. Optimizing Athletic Performance through Advanced Nutrition strategies: Can. AI and Digital Platforms Have a Role in Ultraendurance Sports? *Biol. Sport* **2024**, *41*, 305–313. [CrossRef]
30. Sheikh, H.; Prins, C.; Schrijvers, E. Artificial Intelligence: Definition and Background. In *Research for Policy*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 15–41. [CrossRef]
31. Razno, M. Machine learning text classification model with NLP approach. *Comput. Linguist. Intell. Syst.* **2019**, *2*, 71–73.
32. Gupta, N.; Khatri, K.; Malik, Y.; Lakhani, A.; Kanwal, A.; Aggarwal, S.; Dahuja, A. Exploring Prospects, Hurdles, and Road Ahead for Generative Artificial Intelligence in Orthopedic Education and Training. *BMC Med. Educ.* **2024**, *24*, 1544. [CrossRef]
33. Millington, B.; Naraine, M.L.; Wanless, L.; Safai, P.; Manley, A. Sport and the Promise of Artificial Intelligence: Human and Machine Futures. *Sociol. Sport J.* **2025**, *1*, 1–10.
34. Desai, V. The future of artificial intelligence in sports medicine and return to play. In *Seminars in Musculoskeletal Radiology*; Thieme Medical Publishers Inc.: New York, NY, USA, 2024; Volume 28, pp. 203–212.
35. Lotfi, N.; Madani, M. Evaluating the Qualitative and Quantitative Performance of Generative AI on Knowledge in Sports Medicine: The Case of GPT. In *General Aspects of Applying Generative AI in Higher Education: Opportunities and Challenges*; Springer Nature: Cham, Switzerland, 2024; pp. 103–119.
36. Chemnad, K.; Othman, A. Digital Accessibility in the Era of Artificial Intelligence—Bibliometric Analysis and Systematic Review. *Front. Artif. Intell.* **2024**, *7*, 1349668. [CrossRef]
37. Shuford, J. Contribution of artificial intelligence in improving accessibility for individuals with disabilities. *J. Knowl. Learn. Sci. Technol.* **2023**, *2*, 421–433, ISSN 2959-6386.
38. Kulkarni, M. Digital accessibility: Challenges and opportunities. *IIMB Manag. Rev.* **2019**, *31*, 91–98.
39. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *Syst. Rev.* **2021**, *10*, 89. [CrossRef] [PubMed]
40. UNO per Tutto. Available online: https://unopertutto.unige.net/discovery/search?vid=39GEN_INST:39GEN_VU1 (accessed on 13 March 2025).
41. Sallam, M.; Barakat, M.; Sallam, M. A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review. *Interact. J. Med. Res.* **2023**, *13*, e54704.
42. Masagca, R.C. The AI Coach: A 5-Week AI-Generated Calisthenics Training Program on Health-Related Physical Fitness Components of Untrained Collegiate Students. *J. Hum. Sport Exerc.* **2024**, *20*, 39–56. [CrossRef]

43. Dergaa, I.; Ben Saad, H.; El Omri, A.; Glenn, J.; Clark, C.; Washif, J.; Guelmami, N.; Hammouda, O.; Al-Horani, R.; Reynoso-Sánchez, L.; et al. Using Artificial Intelligence for Exercise Prescription in Personalised Health Promotion: A Critical Evaluation of OpenAI's GPT-4 Model. *Biol. Sport* **2024**, *41*, 221–241. [\[CrossRef\]](#)
44. Dükking, P.; Sperlich, B.; Voigt, L.; Hooren, B.V.; Zanini, M.; Zinner, C. ChatGPT Generated Training Plans for Runners Are Not Rated Optimal by Coaching Experts, but Increase in Quality with Additional Input Information. *J. Sports Sci. Med.* **2024**, *23*, 56–72. [\[CrossRef\]](#)
45. Erol, E.; Arıkan, H. Does ChatGPT Provide Comprehensive and Accurate Information Regarding the Effects, Types and Programming of Core Exercises? *Turk. J. Kinesiol.* **2024**, *10*, 178–182. [\[CrossRef\]](#)
46. Havers, T.; Masur, L.; Isenmann, E.; Geisler, S.; Zinner, C.; Sperlich, B.; Dükking, P. Reproducibility and Quality of Hypertrophy-Related Training Plans Generated by GPT-4 and Google Gemini as Evaluated by Coaching Experts. *Biol. Sport* **2025**, *42*, 289–329. [\[CrossRef\]](#)
47. Philuek, P.; Kusump, S.; Sathianpoonsok, T.; Jansupom, C.; Sawanyawisuth, P.; Sawanyawisuth, K.; Chainarong, A. The Effects of Chat GPT Generated Exercise Program in Healthy Overweight Young Adults. *J. Hum. Sport Exerc.* **2024**, *20*, 169–179.
48. Washif, J.; Pagaduan, J.; James, C.; Dergaa, I.; Beaven, C. Artificial Intelligence in Sport: Exploring the Potential of using ChatGPT in Resistance Training Prescription. *Biol. Sport* **2024**, *41*, 209–220. [\[CrossRef\]](#)
49. Xu, Y.; Liu, Q.; Pang, J.; Zeng, C.; Ma, X.; Li, P.; Ma, L.; Huang, J.; Xie, H. Assessment of Personalized Exercise Prescriptions Issued by ChatGPT 4.0 and Intelligent Health Promotion Systems for Patients with Hypertension Comorbidities Based on the Transtheoretical Model: A Comparative Analysis. *J. Multidiscip. Healthc.* **2024**, *17*, 5063–5078. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Zaleski, A.L.; Berkowsky, R.; Craig, K.J.T.; Pescatello, L.S. Comprehensiveness, Accuracy, and Readability of Exercise Recommendations Provided by an AI-Based Chatbot: Mixed Methods Study. *JMIR Med. Educ.* **2024**, *10*, e51308. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Zhu, W.; Geng, W.; Huang, L.; Qin, X.; Chen, Z.; Yan, H. Who Could and Should Give Exercise Prescription: Physicians, Exercise and Health Scientists, Fitness Trainers, or ChatGPT? *J. Sport Health Sci.* **2024**, *13*, 368–372. [\[CrossRef\]](#)
52. Wang, L.; Chen, X.; Deng, X.; Wen, H.; You, M.; Liu, W.; Li, Q.; Li, J. Prompt Engineering in Consistency and Reliability with the Evidence-Based Guideline for LLMs. *npj Digit. Med.* **2024**, *7*, 41. [\[CrossRef\]](#)
53. Chen, B.; Zhang, Z.; Langrené, N.; Zhu, S. Unleashing the Potential. of Prompt Engineering in Large Language Models: A Comprehensive Review. *arXiv* **2024**, arXiv:2310.14735.
54. Kush, J.C. Integrating Sensor Technologies with Conversational AI: Enhancing Context-Sensitive Interaction Through Real-Time Data Fusion. *Sensors* **2025**, *25*, 249. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Oğul, H. Language of Actions: A Generative Model for Activity Recognition and next Move Prediction from Motion Sensors. *Expert Syst. Appl.* **2025**, *264*, 125947. [\[CrossRef\]](#)
56. Li, Z.; Deldari, S.; Chen, L.; Xue, H.; Salim, F.D. SensorLLM: Aligning Large Language Models with Motion Sensors for Human Activity Recognition. *arXiv* **2024**, arXiv:2410.10624.
57. Casado, A.; González-Mohino, F.; González-Ravé, J.M.; Foster, C. Training Periodization, Methods, Intensity Distribution, and Volume in Highly Trained and Elite Distance Runners: A Systematic Review. *Int. J. Sports Physiol. Perform.* **2022**, *17*, 820–833. [\[CrossRef\]](#)
58. Galán-Rioja, M.Á.; Gonzalez-Ravé, J.M.; González-Mohino, F.; Seiler, S. Training Periodization, Intensity Distribution, and Volume in Trained Cyclists: A Systematic Review. *Int. J. Sports Physiol. Perform.* **2023**, *18*, 112–122. [\[CrossRef\]](#)
59. Mujika, I.; Halson, S.; Burke, L.M.; Balagué, G.; Farrow, D. An Integrated, Multifactorial Approach to Periodization for Optimal Performance in Individual and Team Sports. *Int. J. Sports Physiol. Perform.* **2018**, *13*, 538–561. [\[CrossRef\]](#)
60. Cavazzotto, T.G.; Dantas, D.B.; Queiroga, M.R. ChatGPT and Exercise Prescription: Human vs. Machine or Human plus Machine? *J. Sport Health Sci.* **2024**, *13*, 661–662. [\[CrossRef\]](#)
61. Mishra, N.; Habal, B.G.M.; Garcia, P.S.; Garcia, M.B. Harnessing an AI-Driven Analytics Model to Optimize Training and Treatment in Physical Education for Sports Injury Prevention. In Proceedings of the 2024 8th International Conference on Education and Multimedia Technology, Tokyo, Japan, 22–24 June 2024; ACM: New York, NY, USA, 2024; pp. 309–315.
62. Lanotte, F.; O'Brien, M.K.; Jayaraman, A. AI in Rehabilitation Medicine: Opportunities and Challenges. *Ann. Rehabil. Med.* **2023**, *47*, 444–458. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Biró, A.; Cuesta-Vargas, A.I.; Szilágyi, L. AI-Assisted Fatigue and Stamina Control for Performance Sports on IMU-Generated Multivariate Times Series Datasets. *Sensors* **2023**, *24*, 132. [\[CrossRef\]](#)
64. Rocha-Silva, R.; Rodrigues, M.A.M.; Viana, R.B.; Nakamoto, F.P.; Vancini, R.L.; Andrade, M.S.; Rosemann, T.; Weiss, K.; Knechtle, B.; De Lira, C.A.B. Critical Analysis of Information Provided by ChatGPT on Lactate, Exercise, Fatigue, and Muscle Pain: Current Insights and Future Prospects for Enhancement. *Adv. Physiol. Educ.* **2024**, *48*, 898–903. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Cheng, K.; Guo, Q.; He, Y.; Lu, Y.; Xie, R.; Li, C.; Wu, H. Artificial Intelligence in Sports Medicine: Could GPT-4 Make Human Doctors Obsolete? *Ann. Biomed. Eng.* **2023**, *51*, 1658–1662. [\[CrossRef\]](#)

66. Mekki, Y.M.; Ahmed, O.H.; Powell, D.; Price, A.; Dijkstra, H.P. Games Wide Open to athlete partnership in building artificial intelligence systems. *npj Digit. Med.* **2024**, *7*, 267, Erratum in: *npj Digit. Med.* **2024**, *7*, 291. <https://doi.org/10.1038/s41746-024-01284-5>. [CrossRef] [PubMed] [PubMed Central]
67. Mennella, C.; Maniscalco, U.; De Pietro, G.; Esposito, M. Ethical and Regulatory Challenges of AI Technologies in Healthcare: A Narrative Review. *Heliyon* **2024**, *10*, e26297. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.