Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

CHATGPT AND GENERATIVE AI IN PHYSICAL EDUCATION

Applications Benefits Limitations and Risks for Teachers Students Coaches and Professionals

Kalyd Pierry Ferreira Gonçalves
Faculdade Presidente Antônio Carlos, Brazil
Corresponding author: kalydpierryferreiragoncalves@gmail.com
Orcid ID: https://orcid.org/0009-0009-5612-7995

Abstract

Generative artificial intelligence and large language models can draft lesson plans, explain rules, generate assessment materials, support reflective dialogue, summarize training information, and translate specialist language. Their use in physical education, however, differs from use in text-dominant subjects because learning is embodied, safety-sensitive, socially situated, and dependent on observation of movement and context. This structured integrative narrative review synthesizes direct evidence from physical education, sport coaching, exercise science, and sports medicine together with adjacent evidence from education and AI governance. The review distinguishes four user groups: teachers, students, coaches, and exercise or health professionals. Direct studies indicate promising results for structured feedback, rule learning, content knowledge, tactical instruction, lesson planning, and coach reflection, but the evidence base remains small, heterogeneous, and concentrated in short-term or single-context studies. Evaluations of exercise advice show that outputs can be accurate on many general points while remaining incomplete, inconsistent, poorly individualized, or unsafe when medical screening and contraindications are omitted. The article proposes the Human-Supervised Generative AI Framework for Physical Education and Sport, which classifies tasks by consequence and requires source grounding, data minimization, professional review, disclosure, and outcome monitoring. LLMs are most defensible as drafting, explanation, and reflection tools. They should not independently diagnose injury, prescribe rehabilitation, determine return to play, set high-risk training loads, or make consequential student evaluations. The central recommendation is to evaluate AI use by verified educational or professional outcomes rather than fluency, novelty, or time saved.

Keywords: generative artificial intelligence; ChatGPT; physical education; sport coaching; exercise prescription; large language models; AI literacy; professional judgment

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

1 Introduction

ChatGPT made advanced text generation available through an ordinary conversational interface. Large language models produce responses by estimating probable continuations from patterns learned during training and, in some systems, from tools or retrieved documents. They can reorganize information, generate examples, adapt tone, and respond iteratively. These capabilities make them attractive for educational planning and professional communication, but linguistic plausibility does not establish factual accuracy, pedagogical quality, or clinical safety (Bender et al., 2021; Ji et al., 2023; Kasneci et al., 2023).

Physical education presents a demanding test case. Its aims include motor competence, health knowledge, participation, social learning, confidence, and the ability to make informed decisions about physical activity. Teaching occurs in changing environments where class size, equipment, weather, disability, previous experience, culture, and immediate safety conditions affect what is feasible. Sport coaching and exercise practice add performance targets, workload management, injury history, consent, safeguarding, and professional scope of practice. A system trained primarily on language cannot directly feel fatigue, observe a learner from all relevant angles, or assume responsibility for consequences.

Recent literature has moved beyond speculation. Physical education researchers have examined lesson planning, assessment design, content knowledge, rule learning, tactical instruction, skill feedback, teacher competence, and student adoption. Coaching research has tested whether freely available GPT systems can help analyze referee-player dialogue. Exercise science and sports medicine researchers have evaluated training programs and health recommendations, often finding a mixture of useful general guidance and important omissions (Albaloul et al., 2024; Genç et al., 2026; O'Brien & Prentice, 2025; Rakha, 2026; Zaleski et al., 2024).

The evidence nevertheless remains easy to overstate. A favorable attitude survey does not demonstrate motor learning. A correct answer to a rules question does not validate injury advice. A well-written twelve-week plan for a hypothetical healthy adult does not establish safe personalization for a real person. Model versions, interfaces, system prompts, retrieval features, and safety policies also change quickly, which limits reproducibility and makes product names poor substitutes for documented configurations (Dergaa et al., 2026; Lai, 2026).

This article asks four questions. First, which tasks can LLMs support for teachers, students, coaches, and physical activity professionals? Second, what benefits are supported by direct evidence and what benefits remain plausible but untested? Third, what technical, pedagogical, ethical, privacy, equity, and safety risks follow from current

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

use? Fourth, what governance controls can preserve human responsibility while allowing proportionate experimentation? The review treats generative AI as one component of a sociotechnical practice rather than as an autonomous instructor or coach.

2 Literature Review

2.1 What Large Language Models Do and Do Not Know

Transformer language models represent relationships among tokens and generate sequences conditioned on prompts and prior context (Vaswani et al., 2017). Instruction tuning and human feedback can make responses more useful and conversational, while retrieval can connect generation to selected documents. None of these mechanisms guarantees that a response is true, current, complete, or appropriate for the individual receiving it. Hallucination is not limited to absurd statements; it also includes subtle errors, invented citations, omitted qualifications, and confident generalization beyond the evidence (Bhattacharyya et al., 2023; Ji et al., 2023).

This distinction matters in physical education because much professional knowledge is conditional. A drill that is appropriate for one age group, surface, class size, disability profile, or phase of rehabilitation may be inappropriate in another. The model may generate an apparently coherent plan without access to movement quality, pain behavior, environmental hazards, current workload, or local curriculum requirements. A useful answer therefore depends on the quality of the prompt, the evidence available to the system, and the reviewer's competence to detect missing or unsafe elements.

Embodied knowledge also limits text-only substitution. Motor learning depends on practice, feedback, attention, perception, and task constraints. An LLM can phrase cues or questions, but it does not by itself verify whether the learner executed the movement, interpreted the cue correctly, or improved retention. Multimodal systems may process images or video, yet camera position, occlusion, sampling, labeling, and consent determine what the system can validly infer. Generative fluency should therefore be separated from observation, measurement, and professional judgment.

2.2 Educational Uses for Teachers and Students

For teachers, the most documented low-consequence uses are drafting lesson objectives, activity progressions, adaptations, questioning strategies, rubrics, and parent-facing explanations. Practice-oriented scholarship in physical education emphasizes that teachers must specify curricular outcomes, learner characteristics, available space and equipment, inclusion requirements, and assessment criteria before accepting generated material (Albaloul et al., 2024; Trendowski, 2025a, 2025b). The

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

output functions as a first draft; the teacher remains responsible for alignment, feasibility, safety, and accuracy.

For students, conversational systems can explain rules, generate retrieval-practice questions, provide alternative explanations, simulate oral questioning, and support reflection after practical sessions. A GPT-integrated messaging bot has been studied for badminton rule learning, and an experimental study reported improved handball skills and attitudes when ChatGPT-generated feedback was embedded in a structured reciprocal teaching model with predefined criteria and mandatory human supervision (Lin et al., 2026; Rakha, 2026). These studies support carefully bounded instructional designs, not unrestricted chatbot use.

Generative AI can also help differentiate text, vocabulary, reading level, and language. This may benefit multilingual learners or students who need multiple representations of a concept. The same process can oversimplify technical content or reproduce stereotypes. Inclusive adaptation requires consultation with learners and relevant specialists rather than automatic assumptions about disability, sex, body size, culture, or motivation (UNESCO, 2023).

Assessment is a contested area. Teachers can use LLMs to draft question banks, performance descriptors, scenarios, and feedback templates, but generated criteria may privilege easily verbalized behaviors and neglect embodied learning. Students can use the same systems to produce essays, reflections, and training rationales without demonstrating understanding. Assessment design must therefore make the learning process visible through observation, oral explanation, drafts, decision logs, practical demonstration, and source verification. Detection tools cannot reliably resolve authorship and should not replace fair procedures or evidence of learning (Cotton et al., 2024; Kasneci et al., 2023).

2.3 Uses for Coaches and Sport Organizations

Coaches can use LLMs to draft practice plans, reorganize session notes, generate questions for athlete reflection, summarize policies, and prepare communication for different audiences. O'Brien and Prentice (2025) tested free GPT tools with a predefined framework for analyzing rugby referee-player dialogue. Their study illustrates a defensible pattern: the professional defines the analytic construct, removes identifiable information, constrains the prompt, and checks the output against the source data. It does not show that a general chatbot can independently interpret all coach-athlete communication.

Training design is more consequential. Genç et al. (2026) compared several ChatGPT versions using expert-developed prompts for a hypothetical healthy, moderately trained

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

young adult. The study found meaningful differences among versions and evaluated scientific approach, progression, modification, injury considerations, and practical applicability. The design demonstrates why product version, prompt, population, and rubric must be reported. It also limits inference: a hypothetical resistance-training program cannot validate real-time load adjustment, youth training, para-sport, elite competition, or rehabilitation.

LLMs may assist tactical preparation by converting structured observations into questions, scenarios, or concise summaries. They are not substitutes for validated tracking, video analysis, or domain-specific models. Tactical recommendations depend on opponent behavior, team principles, athlete capabilities, and game state. When an LLM receives summaries rather than raw validated observations, any error in the summary becomes part of the recommendation chain.

2.4 Uses for Exercise and Health Professionals

Exercise professionals may use LLMs to prepare educational handouts, explain established guidelines, draft interview questions, or organize documentation. These tasks can reduce clerical burden if every output is reviewed. Use becomes higher risk when the system selects exercises, determines intensity, interprets symptoms, or gives advice to people with disease, pregnancy, medication use, disability, or recent injury. Professional standards require screening, informed consent, monitoring, and referral decisions that a generic chatbot cannot reliably perform.

Evidence on exercise recommendations is mixed. Zaleski et al. (2024) found that chatbot-generated recommendations were often accurate on included statements but poorly comprehensive, with frequent omissions related to medical clearance and core exercise-prescription elements. Bernal-Utrera et al. (2026) reported poor accuracy for advice to breast cancer survivors compared with clinical practice guidelines. These findings show why a percentage of correct sentences can coexist with an unsafe overall answer: omitted contraindications, follow-up questions, and escalation rules may matter more than fluent general guidance.

Sports medicine raises the strongest boundary. A 2026 scoping review found expanding use across training prescription, rehabilitation, nutrition, injury prevention, mental health, and academic work, while emphasizing uneven accuracy, hallucination, privacy, and governance concerns (Dergaa et al., 2026). An LLM may support communication after a clinician or qualified professional has made a decision. It should not independently diagnose injury, determine return to play, modify rehabilitation in response to pain, or provide emergency triage.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

2.5 Benefits and the Conditions Required to Realize Them

Table 1 Potential Benefits Conditions and Evidence Status

Potential benefit	Mechanism	Condition for defensible use	Evidence status
Faster drafting	Generates plans, prompts, rubrics, and explanations from structured inputs	Teacher or professional checks curriculum, feasibility, safety, and sources	Direct practice-oriented PE literature; limited outcome evidence
More frequent feedback	Produces criterion-referenced wording between human observations	Criteria are predefined and a qualified person verifies the feedback	Direct experimental evidence in one handball context
Accessible explanations	Adjusts language, format, examples, and translation	Meaning is checked; accessibility is not inferred from diagnosis alone	Adjacent education evidence and plausible PE application
Reflective practice	Codes or summarizes transcripts and session notes	Consent, de-identification, a validated framework, and source comparison	Direct exploratory coaching evidence
Content practice	Generates questions and explanations about rules or concepts	Answers are checked against authoritative rules and curricula	Emerging direct studies
Exercise education	Restates established guidelines and prepares handouts	No individualized prescription without screening and professional review	Mixed direct evidence; important omissions reported

Note. Evidence status distinguishes direct physical education or sport evidence from adjacent educational evidence and plausible but untested applications.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Time saving is not an educational outcome. A faster lesson plan has value only if it remains aligned with the curriculum, increases usable preparation time, or improves instruction. Personalization is likewise not established merely because the output includes a learner's name or stated goal. Genuine adaptation requires relevant information, validated decision rules, respect for learner preferences, and feedback from actual performance. The appropriate comparison is with the existing workflow, including the time required to verify and repair the output.

3 Methodology and Materials and Methods

3.1 Review Design and Purpose

This study used a structured integrative narrative review. The design was selected because the literature spans empirical studies, scoping and systematic reviews, professional practice articles, technical scholarship, and institutional guidance. The purpose was interpretive and decision-oriented: to identify defensible uses, separate direct from adjacent evidence, and derive safeguards for different professional roles. The study was not registered as a systematic review and does not claim exhaustive coverage or meta-analytic effect estimates.

3.2 Information Sources and Final Search Date

PubMed was queried directly for health, exercise, rehabilitation, and sports-medicine records. Google Scholar was used as a discovery tool for physical education, coaching, kinesiology, and educational-technology literature. General web search was restricted to academic publishers, DOI records, PubMed pages, and official institutional domains to locate records that were not consistently indexed under physical-education terminology. In this review, "scholarly search results" therefore means candidate records discovered through Google Scholar or domain-restricted scholarly web searching; it does not denote a separate bibliographic database. Scopus and Web of Science were not directly queried.

Publisher platforms, including Frontiers, Springer Nature, Taylor & Francis, Elsevier, JMIR, and journal-hosted pages, were used primarily for full-text retrieval and bibliographic verification. Targeted within-platform searching was used only when a known title, author, or DOI required confirmation. Institutional guidance was obtained from UNESCO, the United States Department of Education, the European Commission, NIST, and the World Health Organization. Searches were updated through September 18, 2026. Searches were limited to English-language sources; no fixed starting-year filter was imposed because foundational technical and health-guidance sources were retained. Priority was given to peer-reviewed sources with a DOI or stable publisher record and to primary institutional documents.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

3.3 Search Strategy

Searches combined technology terms with physical education, sport, coaching, exercise, and professional-risk terms. Core combinations included: (ChatGPT OR "large language model*" OR "generative AI") AND ("physical education" OR kinesiology OR sport OR coach*); the same technology block AND ("exercise prescription" OR rehabilitation OR "sports medicine"); and the same technology block AND (education OR assessment OR integrity OR privacy OR hallucination). Citation chasing was used to locate direct physical education studies referenced by recent scoping and narrative reviews. Appendix A records reproducible search blocks.

The original narrative search was conducted iteratively and did not preserve a single deduplicated export of all discovery results. Initial-hit, duplicate-removal, and exclusion counts therefore cannot be reconstructed retrospectively without creating false precision. The final analytical corpus contains 32 unique sources. During revision, every retained source was rechecked against its publisher, DOI, PubMed, or institutional record, and no duplicate reference remained. Appendix A provides a simplified, non-PRISMA selection record and states explicitly which counts are unavailable.

3.4 Eligibility and Exclusion Criteria

Sources were eligible when they addressed at least one of the following: LLM use in physical education or kinesiology; coaching, athlete development, or sport communication; exercise recommendation or sports medicine; educational affordances and integrity; or AI governance relevant to student, athlete, or health data. Purely promotional material, unsourced opinion, duplicate reports, and general AI studies without a clear connection to education, sport, health, or governance were excluded from the analytic core. Conventional AI for pose estimation, wearable sensing, or prediction was included only when needed to distinguish those systems from LLMs.

3.5 Data Extraction and Thematic Coding

Each source was charted by population, user role, task, model or system where reported, study design, comparator, outcome, principal finding, limitations, and evidence type. Evidence was coded as direct when the study evaluated LLM use in physical education, sport, exercise, or sports medicine; adjacent when it addressed general education or AI reliability with a clear mechanism applicable to the domain; institutional when it stated a governance or health guideline; and author synthesis when this review combined sources into a proposed control. Appendix D provides the complete extraction matrix for the 32-source corpus.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

3.6 Critical Appraisal and Synthesis Rules

Empirical claims were weighted by design, sample, task realism, comparator quality, outcome relevance, and reproducibility. Studies of attitudes or behavioral intention were not treated as evidence of learning or performance. Benchmark correctness was not treated as evidence of safe practice. Reviews were used to map fields and identify sources, but direct studies were preferred for specific claims. Product-specific findings were interpreted cautiously because model versions and interfaces change. The synthesis used conservative language when only conceptual or single-study evidence was available. The confidence labels in Appendix C are structured narrative judgments, not GRADE ratings, statistical scores, or a formally validated evidence hierarchy.

3.7 Single Author Procedure and Limitations

One reviewer performed searching, selection, extraction, and synthesis. No independent duplicate screening or inter-rater reliability is claimed. This limitation increases the risk of missed sources and interpretive bias. To reduce overstatement, the review reports its narrative design, distinguishes evidence types, gives search strings, includes the complete extraction matrix, and presents a narrative confidence appraisal. The resulting framework is a proposed synthesis, not a validated professional standard.

4 Results

4.1 Distribution and Maturity of the Evidence

The direct literature is recent and methodologically diverse. It includes conceptual and practice articles for teachers, content-accuracy tests, technology-acceptance studies, short educational interventions, an exploratory coaching analysis, expert evaluation of generated programs, and scoping reviews. This diversity supports mapping applications but prevents a pooled estimate of effectiveness. The strongest conclusions concern task boundaries and workflow design rather than a universal effect of ChatGPT on learning or performance (Lai, 2026; Zhu et al., 2026).

Evidence is stronger for text-mediated components of physical education than for embodied outcomes. Lesson-plan drafting, rule explanation, question generation, and transcript summarization are close to the models' native capabilities. Movement diagnosis, injury assessment, load prescription, and return-to-play decisions require observations and causal judgment that cannot be inferred safely from conversational fluency. The distance between the model's input and the real-world outcome is therefore a useful risk indicator.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

4.2 Applications by User Group

Table 2 Applications Boundaries and Human Authority by User Group

User	Suitable assistance	Use only with strong controls	Do not delegate autonomously
Teacher	Draft plans, explanations, questions, adaptations, rubrics, communications	Criterion-referenced feedback; assessment support; use of student work or video	Final grades, safeguarding judgments, medical clearance, unsupervised movement diagnosis
Student	Rule practice, concept explanation, quizzes, reflection prompts, language support	Training ideas checked by a teacher or qualified professional	Self-diagnosis, injury treatment, deceptive submission, unsafe challenge generation
Coach	Session drafts, de-identified note summaries, reflective questions, policy explanations	Tactical analysis and load suggestions grounded in validated data	Selection decisions, return to play, punishment, mental-health or medical judgment
Exercise or health professional	Handout drafting, guideline explanation, documentation templates	Screened programs reviewed and monitored by a qualified professional	Diagnosis, emergency triage, rehabilitation progression, medication or contraindication decisions

Note. Authorization depends on developmental stage, professional scope, jurisdiction, institutional policy, data sensitivity, and the consequence of error.

4.3 Instructional Planning Assessment and Feedback

Physical education planning articles provide structured methods for prompting an LLM with grade level, learning standards, class duration, equipment, space, prior learning, inclusion requirements, and assessment evidence (Albaloul et al., 2024; Trendowski, 2025a). This structure reduces generic output but does not prove that the resulting lesson will work. Teachers must check activity density, transition time, supervision, contraindications, and whether adaptations preserve the intended learning outcome.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

The handball experiment by Rakha (2026) offers direct evidence that ChatGPT-based feedback can contribute to skill learning when embedded in reciprocal teaching. Fifty-six university beginners completed a fifteen-session program, and the AI-supported group showed higher post-test scores across assessed skills and positive attitudes. The intervention included predefined criteria and mandatory manual verification. Its results should therefore be attributed to the complete instructional design rather than to unrestricted ChatGPT use. Replication with different teachers, age groups, sports, retention tests, and blinded assessment is needed.

Assessment design can benefit from rapid generation of examples and variants, yet consequential judgments require transparent evidence. A teacher can ask an LLM to propose descriptors and then revise them against curriculum and observed performance. The system should not infer effort, attitude, disability, or potential from sparse text or video. Students should know when AI contributes to feedback, how their data are used, and how to challenge an error.

4.4 Knowledge Learning and AI Literacy

Rule learning and content knowledge are plausible strengths because the task can be grounded in authoritative texts and checked objectively. Studies have examined basketball common content knowledge and a GPT-integrated bot for badminton rules (Albaloul & Killian, 2026; Lin et al., 2026). These systems are most defensible when the official rule set, edition, jurisdiction, and learning objective are specified. A model answering from memory may mix rule versions or invent exceptions.

AI literacy in physical education requires more than prompt technique. Learners need to understand probabilistic generation, source checking, privacy, bias, uncertainty, authorship, and the limits of personalization. They should also practice comparing a generated answer with an official rule, guideline, or observation. Teacher education literature argues that professional competence must include ethical and pedagogical judgment, because technical familiarity alone can increase use without improving decisions (Killian et al., 2023; Yang et al., 2025).

4.5 Coaching Reflection and Communication

Conversation analysis illustrates both promise and risk. LLMs can rapidly segment transcripts, identify repeated phrases, and organize observations under predefined categories. O'Brien and Prentice (2025) used a sport-specific framework and did not upload identifiable or sensitive player information. Their process supports AI-augmented reflection, but thematic labels remain interpretations that should be checked against the transcript and contextual knowledge. Models may miss irony, power relations, nonverbal behavior, or the significance of silence.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Communication drafting can help coaches explain session goals or summarize debriefs, especially when audiences need different levels of detail. The coach must preserve relational responsibility. Automated messages about selection, injury, body composition, discipline, or mental health can create harm even if grammatically polite. Sensitive communication should be written or delivered by a responsible person who understands the athlete and can respond to questions.

4.6 Exercise Prescription and Health Advice

In this review, “exercise and health professionals” means qualified practitioners acting within their legally authorized scope of practice. The category does not imply interchangeable competence among physical education teachers, coaches, personal trainers, strength-and-conditioning professionals, exercise physiologists, athletic trainers, physiotherapists, and physicians. Each profession must follow its own law, credentialing requirements, referral duties, and institutional rules.

The literature distinguishes general educational advice from individualized prescription. A chatbot can restate public guidelines such as the World Health Organization recommendation that adults accumulate regular aerobic and muscle-strengthening activity, but an individual program depends on goals, baseline capacity, medical history, symptoms, medications, preferences, and response to exercise (World Health Organization, 2020). Missing information should trigger questions or referral rather than confident completion.

In the mixed-methods evaluation by Zaleski et al. (2024), generated recommendations were more accurate than comprehensive. Important omissions concerned preparticipation clearance and elements of frequency, intensity, time, and type. This is a consequential failure pattern: the visible advice may appear reasonable while the missing screen or progression rule remains unnoticed. Genç et al. (2026) further showed that model versions differ, so a one-time evaluation cannot authorize indefinite use after system updates.

Professionals should use a risk-tiered rule. Drafting a generic explanation is low risk; proposing a program for a screened healthy adult is moderate risk and requires review; modifying exercise for pain, disease, pregnancy, recent surgery, or suspected injury is high risk and requires qualified assessment. The model’s disclaimer does not transfer accountability or repair an unsafe recommendation.

4.7 Accuracy Hallucination and Reproducibility

LLMs can produce fabricated references and plausible but incorrect claims. Bhattacharyya et al. (2023) documented high rates of fabricated or inaccurate references in generated medical content. Citation risk is especially relevant when

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

students or professionals ask a chatbot to justify training methods. Every cited work must be checked in a scholarly database or publisher record; the model should not be used as the final source-discovery authority.

Reproducibility is weakened by stochastic output, undisclosed system prompts, updates, retrieval sources, and user-specific context. Research reports should identify the model, version or access date, interface, temperature when available, complete prompt set, number of generations, selection procedure, and human editing. Stability should be tested by repeating prompts and examining whether recommendations change in material ways.

Model-update sensitivity limits transfer between studies and practice. A favorable evaluation of one model release, interface, or safety configuration does not authorize use of a later version without reassessment. Substantial provider updates should trigger renewed testing of accuracy, omissions, stability, subgroup performance, and workflow controls.

4.7.1 Text Only and Multimodal Systems

HSGAI-PES applies primarily to generative conversational systems. A text-only model that drafts an explanation and a vision-enabled system that interprets movement video present different validity problems. Video or image analysis requires additional validation of camera placement, pose visibility, occlusion, temporal resolution, movement labels, reference standards, environmental conditions, and subgroup performance. Consent, lawful data handling, and retention controls are also required because images, video, and voice can identify students or athletes.

Risk also depends on the use of the multimodal output. Descriptive feedback reviewed by a qualified teacher may fall within supported judgment, whereas automated selection, grading, injury inference, or return-to-play recommendations are consequential decisions and must not be delegated autonomously. Multimodal capability does not convert a general-purpose model into a validated biomechanical, educational, or clinical assessment system.

4.8 Privacy Consent and Data Governance

Physical education and sport data can include names, age, disability, health conditions, injury history, body composition, video, voice, location, attendance, and performance. Children and athletes may not be free to refuse when a teacher or coach controls participation or selection. Data minimization is therefore essential: do not upload identifiable student, athlete, patient, or client material to a general-purpose service unless an approved institutional arrangement, lawful basis, access control, retention

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

policy, and informed process are in place (European Commission, 2026; UNESCO, 2023).

De-identification is more than deleting a name. A rare injury, team, date, position, voice recording, or combination of traits can re-identify a person. Before use, organizations should define which data classes are prohibited, which tools are approved, whether provider training on inputs is disabled contractually, where data are processed, who can access logs, and how deletion or correction requests are handled. Consent does not excuse unnecessary collection or weak security.

4.9 Bias Equity and Accessibility

Generated advice may encode assumptions about bodies, gender, motivation, ability, resources, or culturally familiar sports. A plan can also presume access to facilities, devices, quiet space, paid subscriptions, or advanced equipment. Unequal access creates a second risk: students with better models, paid features, stronger language skills, or greater prompting experience may receive more useful assistance. Schools should provide common access conditions or design alternatives that do not penalize non-use (UNESCO, 2023; U.S. Department of Education, 2023).

Accessibility should be evaluated with users. Simplifying instructions, translating text, or offering audio can help, but an LLM may remove important nuance or make stereotyped adaptations. A learner with a disability should not be reduced to a diagnostic label. Teachers should ask what support is useful, preserve agency, and consult specialists where required.

4.10 Academic Integrity and Professional Formation

Undisclosed generation can obscure whether a student can reason, plan, or communicate independently. In physical education programs, written rationales often demonstrate knowledge needed for later professional decisions. If an LLM supplies the rationale, a polished submission may conceal weak understanding. Courses should specify permitted uses, require disclosure, and assess the process through oral defense, practical demonstration, source annotation, drafts, or reflective commentary (Cotton et al., 2024).

A complete ban may also be educationally weak when graduates will encounter AI in practice. The more defensible approach is task-specific permission. Students can critique generated plans, identify missing screening questions, compare outputs with guidelines, and document revisions. Such assignments teach verification and professional accountability rather than passive acceptance.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

4.11 Human Supervised Generative AI Framework for Physical Education and Sport

This review proposes the Human-Supervised Generative AI Framework for Physical Education and Sport, abbreviated HSGAI-PES. HSGAI-PES is a decision-support governance framework, not a validated scoring instrument. It is an author synthesis derived from the evidence and governance sources; it has not been psychometrically or empirically validated. The framework supplies a common governance logic, while task authorization remains profession-specific, jurisdiction-specific, and dependent on curriculum, institutional policy, and legally authorized scope of practice.

Table 3 Human Supervised Generative AI Framework for Physical Education and Sport

Risk tier	Examples	Minimum controls	Final authority	Rationale
Tier 1 Drafting	Generic lesson ideas, quiz items, plain-language explanations, formatting	No personal data; source check; human editing	Human user	Low consequence, reversible output, and no direct decision about a person
Tier 2 Supported judgment	Rubrics, criterion feedback, tactical scenarios, screened program drafts	Approved tool; documented prompt; authoritative grounding; competent review; disclosure	Qualified teacher, coach, or professional	Output informs but does not determine a consequential decision
Tier 3 Consequential decision	Grades, selection, workload changes, individualized exercise with health factors	Independent evidence; direct observation; consent; audit trail; right to contest; AI advisory only	Authorized human decision maker	Material effect on rights, opportunity, health, or safety requires accountable human judgment

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Risk tier	Examples	Minimum controls	Final authority	Rationale
Prohibited autonomous use	Diagnosis, emergency triage, return to play, rehabilitation progression, safeguarding, discipline	No autonomous delegation to a general-purpose LLM	Relevant qualified professional or institutional process	Clinical safety and regulated responsibility; due process and safeguarding; or assessment validity

Note. HSGAI-PES is a proposed governance framework rather than a validated scale. Final authorization must follow the relevant profession, jurisdiction, curriculum, institutional policy, and scope of practice.

Tier assignment escalates as consequence severity, personalization, data sensitivity, uncertainty, irreversibility, absence of direct human observation, and vulnerability of the affected person increase. Developmental stage is an explicit modifier: children, adolescents, university students, and preservice teachers differ in consent, safeguarding, AI literacy, independence, and assessment conditions. A rubric drafted for later teacher revision may be Tier 1, whereas direct use in consequential grading is Tier 3. A program draft for a screened healthy adult may be Tier 2, whereas exercise decisions involving disease, pain, pregnancy, recent surgery, or rehabilitation are Tier 3 or outside the permissible role of the user.

The framework uses six steps. Define the educational or professional outcome. Classify the task, affected people, developmental stage, and authorized professional scope. Minimize and protect data. Ground the model in current authoritative material. Require competent human review. Monitor actual outcomes and stop use when errors, inequities, or verification workload exceed the benefit. Competent review means that the reviewer possesses the professional, curricular, clinical, or technical expertise needed to judge the underlying task independently of the AI output. An unqualified person does not provide meaningful oversight merely by approving plausible text.

NIST's govern, map, measure, and manage functions offer organizational support for documenting responsibility and continuous risk management (NIST, 2023). This alignment is contextual: the NIST AI Risk Management Framework is a general governance resource, not evidence that validates HSGAI-PES or any physical-education application.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

5 Discussion

5.1 The Appropriate Unit of Evaluation Is the Workflow

A model cannot be evaluated apart from the task, prompt, source material, reviewer, user population, and consequence. The handball study combined reciprocal teaching, explicit performance criteria, peer activity, and human verification. The coaching study combined a predefined dialogue framework, de-identified data, and expert interpretation. In both cases, the relevant intervention was a supervised workflow rather than a chatbot acting alone (O'Brien & Prentice, 2025; Rakha, 2026).

This workflow perspective explains why contradictory findings can coexist. A model may answer general exercise questions well and still fail as a safe prescriber. It may generate a creative lesson and still overlook transition time or contraindications. Evaluation should therefore include task completion, factual correctness, omissions, time to verify, learner outcome, equity, user understanding, and adverse events. Satisfaction and grammatical quality are secondary measures.

5.2 Augmentation Is Strongest Where Human Observation Remains Central

The most credible applications leave observation, relationship, and accountability with the professional. A teacher observes movement and uses the model to phrase alternative cues. A coach reviews a transcript and uses generated categories to prompt reflection. An exercise professional completes screening and uses the model to draft an explanation. In each case, AI extends a language task while the human retains the embodied and ethical work.

Substitution becomes less defensible as the system moves closer to diagnosis, selection, grading, or load decisions. These decisions require contextual information, calibrated uncertainty, and responsibility for follow-up. General-purpose LLMs are not licensed professionals and do not maintain a duty of care. Their confident tone can conceal this institutional fact.

5.3 Implications for Teacher and Coach Education

Professional programs should teach a verification routine: identify the claim, locate the governing standard or evidence, compare the output, test for omitted conditions, and record the human revision. Prompting should be taught as specification, not magic wording. Students should learn to supply context, request uncertainty, prohibit invented sources, and ask for alternatives, while recognizing that no prompt guarantees correctness.

Programs should also preserve unaided competence. Future teachers and coaches need to plan, observe, explain, and make safety decisions when a system is unavailable

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

or wrong. Assessments should therefore include both responsible AI-supported work and direct demonstrations of independent competence.

5.4 Organizational Governance

Schools, universities, clubs, clinics, and gyms need policies that identify approved tools, prohibited data, permitted tasks, disclosure rules, review requirements, incident reporting, and procurement criteria. The policy should name accountable roles rather than placing responsibility on "the AI." Procurement should examine data retention, training use, security, accessibility, model changes, export, deletion, and audit capabilities (European Commission, 2026; NIST, 2023).

Pilot projects should begin with low-risk, reversible tasks and predefined success measures. A useful pilot compares the existing process with the AI-supported process, including preparation time, correction time, quality, user understanding, and detected errors. Scaling should depend on outcomes, not enthusiasm or vendor claims.

Governance must be localized. National law, data-protection requirements, school curricula, professional regulation, credentialing rules, institutional policy, resources, and sport culture vary internationally. The framework should guide local deliberation rather than override these requirements.

5.5 Research Priorities

Table 4 Research Priorities and Required Outcomes

Priority	Recommended design	Outcome needed
Motor learning	Multi-site controlled studies with retention and transfer tests	Observed skill, retention, safety, and subgroup effects
Teacher workload	Time-and-motion comparison including verification and correction	Net time, quality, cognitive load, and adoption sustainability
Exercise safety	Expert-blinded evaluation across realistic profiles and repeated generations	Omissions, contraindications, escalation, stability, and harm potential
Coaching practice	Longitudinal field studies with athlete perspectives	Decision quality, relationship effects, consent, and performance relevance
Equity and inclusion	Participatory studies across disability, language, sex, age, and resource settings	Differential error, access, usability, and agency

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Priority	Recommended design	Outcome needed
Governance	Prospective audits of approved workflows	Incidents, traceability, contestability, and policy compliance

Note. Future studies should measure task-relevant outcomes and adverse effects rather than relying on satisfaction or intention to use.

Future studies should report exact prompts, model access dates, number of generations, human edits, missing-data behavior, and adverse outputs. They should compare AI-supported practice with a credible alternative, not with no support. Where the outcome is physical, researchers should measure physical performance or behavior rather than infer benefit from intention to use.

5.6 Limitations of This Review

This review has five main limitations. First, the search was structured but not exhaustive, and one reviewer completed the process. Second, initial discovery and exclusion counts cannot be reconstructed because the original narrative search did not preserve a deduplicated export. Third, rapid publication, interface changes, and model updates make conclusions version-sensitive and time-sensitive. Fourth, much of the direct evidence is recent, short-term, based on one or a few studies, and geographically concentrated. Fifth, physical education, coaching, fitness, clinical exercise practice, rehabilitation, and sports medicine have different professional boundaries. A safeguard appropriate in one setting may be insufficient in another. The proposed HSGAI-PES framework requires empirical testing and localization to national law, curriculum, professional regulation, institutional policy, and available resources.

6 Conclusion

ChatGPT and related LLMs can support physical education and sport when the task is primarily linguistic, the source material is verifiable, personal data are protected, and a competent human reviews the result. Useful applications include drafting lesson materials, explaining rules, generating questions, supporting reflection, organizing de-identified notes, and preparing plain-language communication. Emerging studies also suggest value in structured feedback and tactical or content learning, but the evidence is not yet broad enough to justify general claims of improved learning or performance.

The main risks are not limited to obvious hallucinations. Incomplete screening, outdated rules, hidden assumptions, unstable outputs, fabricated references, data exposure, inequitable access, weakened assessment, and misplaced authority can all produce

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

harm. These risks increase when an LLM moves from drafting to consequential judgment.

The practical boundary is clear. Teachers, coaches, and professionals may use LLMs to assist thought and communication; they should not transfer responsibility for embodied observation, safety, grading, selection, diagnosis, rehabilitation, or return to play. Institutions should authorize uses by risk tier, require disclosure and verification, and measure real outcomes. Fluency is an interface property. Educational and professional value must be demonstrated in practice.

Appendix A Search and Selection Record

Table A1 Search Concept Blocks

Concept block	Illustrative search string
Physical education	(ChatGPT OR “large language model*” OR “generative AI”) AND (“physical education” OR kinesiology OR “teacher education”)
Sport and coaching	(ChatGPT OR “large language model*” OR “generative AI”) AND (sport OR coach* OR athlete* OR tactical OR “performance feedback”)
Exercise and health	(ChatGPT OR “large language model*” OR “generative AI”) AND (“exercise prescription” OR rehabilitation OR “sports medicine” OR “physical activity”)
Education and governance	(ChatGPT OR “generative AI”) AND (assessment OR integrity OR privacy OR equity OR hallucination OR governance)

Note. Search syntax was adapted to the interface. Quotation marks identify phrases; the asterisk indicates truncation where supported.

Table A2 Simplified Source Selection Record

Stage	Record
Discovery	PubMed direct search; Google Scholar discovery; domain-restricted scholarly web search; citation chasing; official institutional repositories
Full-text retrieval	Publisher platforms and DOI or PubMed records were used primarily for retrieval and metadata verification
Initial records and duplicate removal	Not reconstructable because the iterative narrative search did not preserve a single deduplicated export

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Stage	Record
Eligibility confirmation	Retained sources were rechecked against eligibility criteria and authoritative bibliographic records during revision
Final analytical corpus	32 unique sources
Automated screening	None

Note. This is a transparent narrative-review record, not a PRISMA flow diagram. No unrecorded numerical counts were retrospectively invented.

Final search date: September 18, 2026. The review did not use automated screening, did not register a protocol, and does not claim systematic-review or PRISMA compliance. English-language sources were included; no fixed starting-year filter was used because foundational sources were eligible.

Appendix B Selected Source Characteristics

Table B1 Selected Source Characteristics

Source	Context and design	Contribution to synthesis	Main limitation
Albaloul et al. 2024	Practice-focused PE article	Teacher planning, instruction, and assessment uses	Does not establish learning effects
Killian et al. 2023	Conceptual HPE teacher education reflection	Professional and teacher-education implications	Early conceptual evidence
O'Brien and Prentice 2025	Exploratory GPT-assisted dialogue analysis	Coach reflection, prompting, de-identification	Specific framework and dataset
Zaleski et al. 2024	Mixed-methods evaluation of exercise recommendations	Accuracy can coexist with poor completeness	Model and prompt dependent
Genç et al. 2026	Expert evaluation across ChatGPT versions	Training-plan quality varies by version and criterion	Hypothetical healthy adult

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Source	Context and design	Contribution to synthesis	Main limitation
Rakha 2026	Randomized pre-post handball study, n = 56	Direct skill and attitude evidence under supervision	Single course and short-term context
Albaloul and Killian 2026	Basketball content-knowledge evaluation	Tests domain knowledge rather than general fluency	Knowledge score does not equal teaching quality
Lin et al. 2026	GPT-integrated bot for badminton rules	Grounded conversational support for rule learning	Narrow content domain
Lai 2026	Scoping review with SWOT synthesis	Maps sport and PE opportunities and threats	Dependent on emerging heterogeneous studies
Zhu et al. 2026	Narrative review and conceptual framework	Maps PE and health applications and interdisciplinary learning	Broad scope and mixed evidence
Dergaa et al. 2026	Sports medicine scoping review	Accuracy, ethics, hallucination, governance	Rapidly changing evidence base
UNESCO 2023	Global institutional guidance	Human-centered use, privacy, equity, age appropriateness	General rather than PE-specific

Note. This compact table highlights the most decision-relevant direct and institutional sources. Appendix D reports the complete 32-source extraction matrix.

Appendix C Narrative Confidence Appraisal

Table C1 Narrative Confidence Appraisal

Evidence claim	Confidence	Reason for rating
LLMs can accelerate first-draft lesson materials	Moderate	Direct practice literature and strong task-capability fit; limited measured workload outcomes
Structured AI-supported feedback can improve motor skill learning	Low to moderate	Positive randomized study, but one setting and bundled intervention

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Evidence claim	Confidence	Reason for rating
LLMs can support rule and content learning	Low to moderate	Emerging direct studies; narrow domains and changing systems
LLMs can support coach reflection from transcripts	Low	Exploratory evidence with predefined framework; context interpretation remains human
Generic LLM exercise advice is safe without professional review	Very low and contradicted	Direct evaluations report incomplete or inaccurate recommendations in consequential contexts
AI use improves teacher or coach professional competence	Low	Associations and acceptance evidence do not establish durable competence
Privacy and governance controls are necessary	High	Consistent institutional guidance and sensitivity of student, athlete, and health data
Autonomous diagnosis, rehabilitation, or return-to-play decisions are appropriate	Unsupported	No validating evidence and consequences require qualified professional responsibility

Note. Confidence categories are structured single-author judgments based on study design, sample, task realism, comparator quality, outcome relevance, and reproducibility. They are not GRADE ratings, statistical scores, or a validated evidence-grading instrument.

Appendix D Complete Extraction Matrix

Table D1 Complete Extraction Matrix for the Analytical Corpus

Source	Evidence class	Domain or task	Design or source type	Principal limitation
Albaloul & Killian 2026	Direct PE	Basketball content knowledge	Accuracy evaluation	Tests domain knowledge; not teaching quality
Albaloul et al. 2024	Direct PE practice	Planning instruction assessment	Practice article	No measured learning outcome
Bender et al. 2021	Adjacent technical ethics	LLM limitations and bias	Conceptual analysis	Not PE specific

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Source	Evidence class	Domain or task	Design or source type	Principal limitation
Bernal-Utrera et al. 2026	Direct health	Exercise advice for breast cancer survivors	Guideline comparison	One clinical population and model context
Bhattacharyya et al. 2023	Adjacent medical reliability	Generated reference accuracy	Empirical reference audit	Medical content rather than PE
Cotton et al. 2024	Adjacent education	Academic integrity and assessment	Conceptual and practice analysis	Higher education focus
Dergaa et al. 2026	Direct sports medicine	Applications accuracy ethics	Scoping review	Rapidly evolving and heterogeneous evidence
European Commission 2026	Institutional	Ethical AI and data use in education	Guidance	Requires national implementation
Genç et al. 2026	Direct exercise science	Training program design	Expert mixed-method evaluation	Hypothetical healthy adult
Ha et al. 2024	Direct kinesiology	Opportunities and professional considerations	Conceptual practice article	Predominantly prospective claims
Ji et al. 2023	Adjacent technical	Hallucination taxonomy	Survey review	Not domain specific
Kasneci et al. 2023	Adjacent education	Educational opportunities and risks	Interdisciplinary review	General education evidence
Killian et al. 2023	Direct HPE teacher education	Professional and educational implications	Conceptual reflection	Early evidence
Lai 2026	Direct sport and PE	Strengths weaknesses opportunities threats	Scoping review	Heterogeneous included evidence

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Source	Evidence class	Domain or task	Design or source type	Principal limitation
Lee & Kim 2025	Direct PE practice	Professionalism and implementation	Practice article	Limited outcome evidence
Lin et al. 2026	Direct PE learning	Badminton rule learning	GPT-integrated learning study	Narrow rule domain
Lo 2023	Adjacent education	Educational impact of ChatGPT	Rapid review	Early and cross-domain literature
Miao et al. 2025	Direct sport pedagogy	AI-assisted tactical instruction	Crossover experiment	Male college sample and bounded task
NIST 2023	Institutional governance	AI risk management	General framework	Does not validate PE applications
O'Brien & Prentice 2025	Direct coaching	Dialogue analysis and reflective practice	Exploratory empirical study	Specific framework and dataset
Rakha 2026	Direct PE motor learning	Handball feedback within reciprocal teaching	Randomized pre-post study	Single course and bundled intervention
Tlili et al. 2023	Adjacent education	Chatbot affordances and risks	Case-based analysis	Not PE specific
Trendowski 2025a	Direct PE practice	Lesson-plan generation checklist	Practice framework	No controlled outcome evaluation
Trendowski 2025b	Direct PE practice	Five generative AI applications	Practice article	Conceptual implementation guidance
US Department of Education 2023	Institutional	Human-centered AI in education	Policy report	US-specific policy context
UNESCO 2023	Institutional	Generative AI governance in education	Global guidance	Not PE-specific

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Source	Evidence class	Domain or task	Design or source type	Principal limitation
Vaswani et al. 2017	Foundational technical	Transformer architecture	Technical paper	Does not evaluate current educational use
World Health Organization 2020	Institutional health	Physical activity guidance	Global guideline	General guidance rather than individualized prescription
Wu et al. 2025	Direct PE adoption	Behavioral intention and use	Technology-acceptance study	Intention does not establish learning
Yang et al. 2025	Direct PE teacher education	Self-regulation TPACK and ethics	Empirical association study	Associations do not establish causal competence
Zaleski et al. 2024	Direct exercise advice	Comprehensiveness accuracy readability	Mixed-method evaluation	Model and prompt dependent
Zhu et al. 2026	Direct PE and health	Applications and thematic learning	Narrative review and framework	Broad scope and mixed evidence

Note. Direct evidence evaluates generative AI within physical education, sport, exercise, kinesiology, or sports medicine. Adjacent evidence addresses education or technical reliability with a transferable mechanism. Institutional sources provide governance or health guidance. The matrix was compiled by one author; no inter-rater reliability is claimed.

References

Albaloul, O., & Killian, C. M. (2026). Evaluating ChatGPT accuracy in answering a basketball common content knowledge test in physical education. Education and Information Technologies, 31, 1793–1808. https://doi.org/10.1007/s10639-025-13873-2

Albaloul, O., Marttinen, R., & Killian, C. (2024). Unlocking educational potential: How physical education teachers can thoughtfully benefit from using ChatGPT for planning, instruction, and assessment. Journal of Physical Education, Recreation & Dance, 95, 32–41. https://doi.org/10.1080/07303084.2024.2378776

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 610–623. https://doi.org/10.1145/3442188.3445922

Bernal-Utrera, C., Bravo-Vázquez, A., Montero-Bancalero, F. J., Suárez-Vega, A., Casuso-Holgado, M. J., & Anarte-Lazo, E. (2026). Evaluation of ChatGPT accuracy and reliability in answering questions about exercise recommendations for breast cancer survivors. Physiotherapy, 130, 101838. https://doi.org/10.1016/j.physio.2025.101838

Bhattacharyya, M., Miller, V. M., Bhattacharyya, D., & Miller, L. E. (2023). High rates of fabricated and inaccurate references in ChatGPT-generated medical content. Cureus, 15, e39238. https://doi.org/10.7759/cureus.39238

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61, 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dergaa, I., Dergaa, M. A., Razzak, M., Ceylan, H. İ., Stefanica, V., Muntean, R. I., & Guelmami, N. (2026). Generative artificial intelligence and large language models in sports medicine: A scoping review of applications, accuracy, and ethical implications. Frontiers in Public Health, 14, 1843535. https://doi.org/10.3389/fpubh.2026.1843535

European Commission. (2026). Guidelines on the ethical use of artificial intelligence and data in teaching and learning for educators. Publications Office of the European Union.

Genç, A., Aydın, G. R., Kasap, M., Özkan, A., Rakhymzhanov, A., Gürcan, H. H., Güllü, S., & Demirhan, B. (2026). Comparative evaluation of ChatGPT versions in training program design: Scientific approach, accuracy, and practical applicability. BMC Sports Science, Medicine and Rehabilitation, 18, 19. https://doi.org/10.1186/s13102-025-01409-7

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

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), Article 248. https://doi.org/10.1145/3571730

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

Killian, C. M., Marttinen, R., Howley, D., Sargent, J., & Jones, E. M. (2023). Knock, knock…who's there? ChatGPT and artificial intelligence-powered large language models: Reflections on potential impacts within health and physical education teacher education. Journal of Teaching in Physical Education, 42, 385–389. https://doi.org/10.1123/jtpe.2023-0058

Lai, C. Y. (2026). The role of ChatGPT in sport and physical education: A scoping review using the SWOT analysis framework. Journal of Hospitality, Leisure, Sport & Tourism Education, 38, 100602. https://doi.org/10.1016/j.jhlste.2026.100602

Lee, Y., & Kim, D. (2025). Evolving professionalism in the AI era: Implementing generative AI in physical education. Journal of Physical Education, Recreation & Dance, 96, 8–13. https://doi.org/10.1080/07303084.2024.2444586

Lin, K. C., Hung, H. C., & Chen, N. S. (2026). Enhancing badminton rule learning through a GPT-integrated LINE bot. Journal of Computer Assisted Learning, 42. https://doi.org/10.1002/jcal.70231

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13, 410. https://doi.org/10.3390/educsci13040410

Miao, C., Zhang, S., & Liu, K. (2025). Comparing AI-assisted and traditional tactical instruction: A crossover experimental study among male college students. Acta Psychologica, 259, 105431. https://doi.org/10.1016/j.actpsy.2025.105431

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). https://doi.org/10.6028/NIST.AI.100-1

O'Brien, K. A., & Prentice, S. (2025). Assessing the practicality of using freely available AI-based GPT tools for coach learning and athlete development. Frontiers in Sports and Active Living, 7, 1627685. https://doi.org/10.3389/fspor.2025.1627685

Rakha, A. H. A. H. (2026). Digital feedback via free ChatGPT within the reciprocal teaching style: Improving fundamental handball skills and students' attitudes among university beginners. Frontiers in Sports and Active Living, 8, 1772502. https://doi.org/10.3389/fspor.2026.1772502

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x

Trendowski, T. (2025a). A checklist for using generative artificial intelligence to create lesson plans in K–12 physical education. Journal of Physical Education, Recreation & Dance, 96, 15–27. https://doi.org/10.1080/07303084.2024.2437988

Trendowski, T. (2025b). Five ways to use generative artificial intelligence in physical education. Strategies, 38, 3–12. https://doi.org/10.1080/08924562.2024.2444202

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 12/08/2026 - **Accepted:** 24/09/2026
Vol: 01.05
DOI: 10.66104/027mff80
Pages: 1-29

United States Department of Education, Office of Educational Technology. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations.

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

World Health Organization. (2020). WHO guidelines on physical activity and sedentary behaviour. World Health Organization.

Wu, Q. J., Li, S. S., Xin, S., Hou, Q., & Li, P. (2025). A study on students' behavioural intention and use behaviour of artificial intelligence-generated content in physical education: Employing an extended unified theory of acceptance and use of technology model. Journal of Hospitality, Leisure, Sport & Tourism Education, 36, 100547. https://doi.org/10.1016/j.jhlste.2025.100547

Yang, P., Liu, Y., & Li, Y. (2025). The mediating role of self-regulation in fostering intelligent-TPACK and ethics in physical education teacher education students. Scientific Reports, 15, 34126. https://doi.org/10.1038/s41598-025-21795-9

Zaleski, A. L., Berkowsky, R., Craig, K. J. T., & Pescatello, L. S. (2024). Comprehensiveness, accuracy, and readability of exercise recommendations provided by an AI-based chatbot: Mixed methods study. JMIR Medical Education, 10, e51308. https://doi.org/10.2196/51308

Zhu, G., Dai, X., Jiang, S., Wang, Z., Liang, M., Pan, Y., & Tao, Y. (2026). Generative AI in physical education and health: A narrative review and conceptual framework for interdisciplinary thematic learning. Frontiers in Public Health, 14, 1842782. https://doi.org/10.3389/fpubh.2026.1842782