Reflections on and Framework for AI-Enabled Evaluation in Physical Education

Chuyang Wu, Aoen Wen*

Guangzhou College of Applied Science and Technology, Zhaoqing, Guangdong, China
Corresponding Author

Abstract: Amidst the ongoing digital transformation and the imperative for high-quality education, artificial intelligence (AI) technology presents unprecedented opportunities to overcome the inherent limitations of traditional physical education evaluation. This study employs a systematic review and logical analysis to examine the significance, current challenges, and strategic responses associated with AI integration in physical education evaluation. It argues that AI enables a fundamental shift from static, one-dimensional, and outcome-focused evaluations to dynamic, multi-faceted, and development-oriented approaches. This transition enhances evaluation accuracy, facilitates personalized feedback, optimizes instructional decision-making, and promotes collaborative engagement among diverse stakeholders. However, practical application reveals several challenges: technologically, issues such as fragmented data silos, opaque algorithms, and insufficient adaptability to complex educational environments persist; within the educational domain, risks include overreliance on technology, diminishing educator expertise, and the erosion of the evaluative process's humanistic qualities; ethically and governance-wise, concerns center on privacy and security risks, algorithmic bias, and the lack of robust accountability mechanisms. To address these challenges, the study proposes an integrated pathway that synergizes technology, education, and ethics by advancing data standardization and explainable AI, establishing a human-AI collaborative evaluation model, and strengthening data security alongside algorithmic governance. This framework aims to cultivate a scientific, equitable, and human-centered evaluation system for physical education.

Keywords: Artificial Intelligence; Physical Education; Human-AI Collaboration; Paradigm Shift

1. Introduction

Educational evaluation steers the direction of educational practice, and its underlying philosophy and methodology critically influence the realization of educational objectives and the holistic development of students. Historically, physical education evaluation in China has been profoundly shaped by traditional evaluation paradigms, which tend to emphasize skills over literacy, outcomes over processes, uniformity over individuality, and quantitative measures over qualitative insights. The evaluation content predominantly centers on standardized, quantifiable athletic skills and physical health test data; the evaluators are primarily teachers, with students positioned largely as passive recipients; and the evaluation methods rely heavily on summative examinations, neglecting the nuanced and developmental dimensions such as learning processes, attitudes, teamwork, and the cultivation of healthy behaviors [1]. Although this evaluation model facilitates management and comparability, it falls short of aligning with contemporary physical education philosophies that prioritize "health first", "educating through sports", and "individual development", thereby, to some extent, constraining the full realization of the educational potential inherent in physical education [2].

In recent years, AI technologies—such as big data analytics, machine learning, computer vision, and natural language processing—have advanced rapidly, permeating various sectors of society with unprecedented depth and scope [3]. Within the realm of education, AI is transitioning from a supplementary teaching tool to a central catalyst that is reshaping the educational landscape. Its applications in learning analytics, adaptive instruction, and automated evaluation demonstrate remarkable

potential [4], offering a unique and timely opportunity to systematically reform the evaluation of physical education [5]. Specifically, AI harnesses multimodal sensing technologies to unobtrusively and continuously collect diverse data during students' physical activities—including physiological metrics, behavioral patterns, posture, and even emotional states. Through sophisticated algorithms, this rich data can be thoroughly analyzed to deliver multidimensional, process-focused, and personalized evaluations of motor skills, engagement levels, physical and mental well-being, and teamwork abilities. Such advancements promise to shift physical education evaluation from a traditional, static "post-event judgment" to a dynamic, continual process of optimization and developmental guidance [6].

However, the integration of AI into physical education evaluation is not without significant challenges. Technical limitations, such as fragmented data ecosystems and the opacity of AI algorithms, impede seamless implementation. Educationally, there is the risk of overdependence on technology, which could diminish educators' professional judgment and undermine the humanistic essence of evaluation. Ethically, concerns regarding data privacy, algorithmic bias, and insufficient accountability frameworks raise serious questions about fairness and trustworthiness [7]. Current research often isolates technical achievements in specific sports skill evaluations or broadly contemplates AI's potential and risks in education, falling short of a comprehensive, interdisciplinary approach that addresses the unique characteristics of physical education. There is a pressing need for a systematic analysis that synthesizes technological capabilities, pedagogical requirements, and ethical considerations [8].

Therefore, this study aims to respond to three fundamental questions: (1) What are the profound and indispensable values that AI technology brings to the reform of physical education evaluation? (2) What specific technical, practical, and ethical challenges arise in transitioning AI from theoretical promise to widespread application? (3) How can a collaborative, multi-stakeholder framework be constructed to effectively address these challenges, ensuring that AI-enabled physical education evaluation develops in a manner that

is scientifically rigorous, equitable, and ethically sound? By thoroughly examining these issues, this research seeks to provide policymakers, educators, and technology developers with well-founded theoretical insights and actionable guidance, thereby promoting the robust construction and sustainable innovation of physical education evaluation systems in the era of AI.

2. The Multifaceted Significance of AI Empowerment in Physical Education Evaluation

The integration of AI transcends a mere technological enhancement of evaluation tools; it catalyzes profound transformations across multiple dimensions of physical education evaluation—encompassing philosophy, methodology, stakeholders, and function—ushering in a paradigm shift with far-reaching innovative implications.

2.1 Perspective on Evaluation Philosophy

Traditional evaluations primarily serve to distinguish performance levels and assign rankings. Through the continuous collection and analysis of students' physical activity data over an entire semester or academic period, AI can generate personalized digital growth profiles. These profiles not only document mastery of critical skills but also visually map progress trajectories, effort intensity, strengths, and areas needing improvement. This fundamentally shifts the focus of evaluation from cross-sectional, summative comparisons to longitudinal, process-oriented "self-relative" evaluations, thereby fully embodying the philosophy of developmental evaluation. The purpose of evaluation evolves from "labeling" to "fostering growth", offering tailored exercise recommendations and goal-setting for each student, which in turn stimulates intrinsic motivation and transforms compliance-based participation ("I have to practice") into autonomous engagement ("I want to practice") [9].

2.2 Perspective on Evaluation Content

Constrained by technological limitations, traditional evaluations struggle to effectively appraise core physical education literacies such as motor competence, health behaviors, and sportsmanship. AI vastly enriches and deepens evaluative content. For example, video analysis

can not only assess the technical correctness of a basketball shot but also interpret decision-making through running patterns and passing timing within game footage. Wearable devices monitor heart rate variability and sleep quality before and after exercise, indirectly evaluating the positive effects of physical activity on physical and mental health. Furthermore, by analyzing communication data and collaborative behavior patterns in team sports, AI facilitates quantifiable observation of attributes like teamwork spirit and responsibility. By integrating such heterogeneous data sources, AI constructs comprehensive, multidimensional digital profiles that encompass knowledge, skills, emotions, attitudes, and values, thereby achieving holistic and synthesized evaluation of students' physical literacy.

2.3 Perspective on Evaluation Methods

Traditional evaluation by physical education instructors heavily relies on personal experience and instantaneous observations, which are inherently subjective and difficult to replicate. AI introduces data-driven, objective, and quantitative analyses. For motor skills, computer vision technology can deliver millimeter-level precision in joint angle measurements and motion trajectory analysis, yielding evaluations far more consistent and accurate than human perception. Regarding physical exertion, heart rate monitoring scientifically gauges whether training intensity is appropriate, effectively preventing both overtraining and undertraining. This addresses the limitations of subjective human judgment, endowing evaluation outcomes with greater scientific validity and credibility. Moreover, AI's capacity for real-time analytical feedback is revolutionary. Upon completion of a movement, the system instantly displays deviations from the standard model and offers visualized correction guidance, facilitating a seamless cycle of "immediate learning, evaluation, and feedback", thereby dramatically enhancing learning efficiency.

2.4 Perspective on Evaluation Stakeholders

AI systems can function as impartial and efficient platforms that integrate evaluative inputs from diverse stakeholders. Students can engage in self-reflection and documentation through the system; peers may conduct more objective mutual evaluations using clearly defined AI-generated criteria; parents gain authorized access to insights regarding their child's physical condition and activity participation, enabling constructive familial feedback. AI undertakes the aggregation, analysis, and visualization of these multifaceted data streams. This disrupts the traditional paradigm wherein teachers hold exclusive evaluative authority, fostering a democratized community of evaluation involving students, peers, educators, parents, and AI itself. Such a model not only enhances objectivity but also cultivates students' metacognitive abilities and social competence.

2.5 Perspective on Instructional Support

In conventional teaching settings, it is challenging for educators to comprehensively grasp the individualized learning status of dozens of students. AI serves as an invaluable instructional aid. Through analytic dashboards, teachers gain immediate, holistic insights—for instance, the collective proficiency of the class regarding a particular technical skill; students at potential risk of sports injuries; individuals exhibiting exceptional talent who require additional focus; and the most effective organizational formats that foster collaborative learning. This liberation from onerous data compilation and rudimentary judgments allows instructors to dedicate greater attention to personalized coaching, optimized lesson design, and meaningful teacher-student interactions. Consequently, teaching can evolve from experience-based, generalized "broad-stroke" approaches toward data-informed, precise, and differentiated pedagogy, substantially enhancing instructional efficacy and professional growth.

3. Contemporary Challenges in AI-Enabled Physical Education Evaluation

Despite its promising potential, the deep integration and application of AI in physical education evaluation confront formidable challenges across three critical dimensions: technological limitations, educational practice, and ethical governance [10].

3.1 Technological Bottlenecks

At the data level, an ideal intelligent evaluation system demands comprehensive, high-quality, and coherent data streams. However, in reality, campus sports data is exceedingly fragmented: physical fitness test results reside in one system; classroom video recordings are stored locally;

data from students' wearable smartwatches belong to proprietary commercial platforms; and psychological survey outcomes exist as isolated files. The absence of unified data standards and exchange protocols has led to a proliferation of data silos, impeding the formation of a holistic "digital twin". Furthermore, the inherent complexity of the physical education environment—characterized by heterogeneous equipment precision and variable student adherence to usage protocols—generates abundant "noise data", which adversely affects analytical accuracy.

From an algorithmic perspective, although current deep learning models demonstrate exceptional performance, their internal decision-making processes remain opaque. When the system labels a student's movement as "unsatisfactory" or predicts a "high risk of injury", teachers and students typically receive only the outcome, without insight into the specific features considered or the weighting assigned. This lack of explainability critically undermines the educational credibility of the evaluation and hampers the detection and correction of potential algorithmic errors. More gravely, algorithmic bias may be embedded: if training data predominantly represent urban male students or athletic elites, evaluations of rural students, females, or average performers could systematically skew, exacerbating educational inequities.

Regarding application scenarios, physical education encompasses both closed skills performed in controlled environments and open skills requiring real-time tactical decisions and team collaboration. Existing AI evaluation technologies primarily target closed, standardized settings, faltering when applied to dynamic, unpredictable open-skill scenarios or subjective domains such as sportsmanship and the joy of movement. Consequently, significant gaps remain in the adaptability and contextual relevance of current AI methodologies.

3.2 Challenges in Educational Practice

The erosion of teacher agency and the risk to their professional competence represent significant concerns. Excessive reliance on AI systems may engender a "de-skilling" phenomenon wherein educators gradually lose their acute observational acuity, intuitive judgment founded on experience, and the capacity for qualitative evaluation through interpersonal dialogue. Prolonged dependence on algorithmic evaluations risks reducing teachers to mere operators and interpreters of systems, thereby depriving them of their essential protagonism and creative autonomy in the evaluative process.

The "de-contextualization" of evaluation and the absence of humanistic care pose profound challenges. AI-driven evaluations are data-centric, yet education—particularly physical education—is steeped in emotional exchange, the forging of willpower, and the nurturing of character. How might a student's selfless sacrifice for team success, which may negatively affect individual performance metrics, be quantified in an algorithm? How can a physically disadvantaged student's arduous efforts and personal progress be fully represented through purely objective data? The relentless pursuit of quantification risks divorcing evaluation from its socio-cultural context and the lived experiences of individuals, thereby neglecting the intangible dimensions of emotion, attitude, and values inherent in sports, resulting in a cold, one-dimensional evaluation.

A crisis emerges concerning students' developmental autonomy. Continuous and comprehensive digital monitoring of athletic performance may impose an invisible pressure and sense of surveillance. Students might engage in utilitarian training aimed solely at optimizing data indicators, forfeiting the intrinsic joy of physical activity. More fundamentally, this paradigm could shape a "data-driven self" awareness, whereby students define their athletic identity and bodily worth through algorithmic feedback, potentially undermining their self-concept and intrinsic motivation for sport.

3.3 Ethical and Governance Dilemmas

A critical crisis looms over the privacy and security of data in physical education evaluation. The information collected—comprising highly sensitive biometric details such as precise body measurements, real-time physiological data, and accurate location and behavioral trajectories—carries immense vulnerability. Should such biometric data be leaked or maliciously exploited, the consequences could be catastrophic. Presently, many educational institutions and ed-tech companies lack stringent regulations governing data collection, storage, transmission, usage, and destruction, resulting in fragile security safeguards and substantial risks

of privacy breaches.

Algorithmic bias further exacerbates structural inequities within education. The fairness of algorithms is inextricably tied to the quality and representativeness of the training data. When datasets reflect historical or societal prejudices, the algorithms inevitably learn and amplify these biases. For instance, a model trained predominantly on elite athlete data might systematically assign lower scores to average students or those with disabilities—a practice that is inherently unjust, diminishes students' athletic self-confidence, and perpetuates existing inequalities.

Moreover, the ambiguity surrounding responsibility and the absence of robust accountability mechanisms compound these issues. When intelligent evaluation systems generate erroneous judgments that result in adverse consequences, the locus of liability remains undefined. Legal frameworks and regulatory structures in this domain are either nascent or ambiguous, making it challenging to enforce accountability. This deficiency, in turn, dampens stakeholders' enthusiasm and willingness to adopt such technologies on a broad scale.

4. Advancing a Collaborative Approach to AI-Enabled Physical Education Evaluation

4.1 Technological Pathways

Establishing unified educational data standards to foster interoperability is imperative. Led by educational authorities in collaboration with technical experts and specialists in physical education, national or industry-wide standards should be formulated to govern the collection, encoding, storage, and exchange of physical education data. Mandatory compliance with open interface protocols for all school-based sports information systems and smart devices will dismantle data silos, laying the crucial groundwork for an integrated intelligent evaluation platform.

Developing explainable and equitable AI models must be prioritized. In algorithmic design, the adoption and refinement of explainable artificial intelligence (XAI) techniques should ensure that evaluation outcomes are accompanied by transparent, comprehensible rationales. Concurrently, fairness must be ingrained as a fundamental design principle throughout the algorithm lifecycle—incorporating bias detection and mitigation strategies, training on diverse and representative datasets, and institutionalizing routine audits of algorithmic equity.

Advancing specialized technologies tailored to the complexities of sports scenarios is essential. Interdisciplinary research should be galvanized to surmount challenges in group behavior recognition, tactical intent interpretation, and affective as well as intentional computing, thereby enhancing AI's capabilities in open-skill evaluations and holistic competency evaluations. Furthermore, the development of lightweight, cost-effective solutions is vital to accommodate the varying economic realities of educational institutions.

4.2 Educational Pathways

Clarifying the "augmentation, not replacement" paradigm in human-AI collaboration is imperative. It must be explicitly articulated in both philosophy and policy that AI serves as an enhancement tool for educators rather than a substitute. Its intrinsic value lies in processing vast datasets, automating repetitive tasks, and offering objective references; meanwhile, contextual interpretation, value judgments, emotional engagement, and ultimate educational decision-making must remain firmly within the teacher's purview.

Establishing a hybrid evaluation workflow is essential, characterized by a standardized process of "AI preliminary analysis—teacher review and contextual adjustment—dialogue and feedback between teacher and student." For instance, AI might generate detailed motion analysis and physiological trend reports, which educators then enrich and refine through classroom observations and student interviews. The final evaluation thus emerges as a synthesis of data-driven intelligence and pedagogical insight, culminating in warm, meaningful communication with the learner.

Strengthening teachers' competencies in intelligent education is paramount. Incorporating AI literacy into both pre-service training and ongoing professional development is critical. Curriculum should encompass understanding the fundamental principles and limitations of intelligent evaluation, cultivating the ability to critically interpret AI-generated analytics, mastering strategies to harmonize AI outputs with human judgment in evaluations, and enhancing awareness of data ethics. This will

empower educators to become adept navigators and discerning critics of intelligent technologies.

4.3 Ethical and Governance Pathways

Establishing comprehensive security protocols throughout the entire lifecycle of educational data is paramount, with particular emphasis on the stringent protection of sensitive biometric information. Adopting the "privacy by design" principle mandates embedding privacy safeguards from the inception of system development. Measures such as data minimization, purpose limitation, encrypted storage, anonymization, and tiered access controls should be rigorously implemented. Clear regulations regarding data ownership and usage rights must be codified to robustly safeguard students' data subject rights.

Instituting algorithmic impact evaluation and auditing frameworks is essential. All intelligent evaluation systems deployed within educational settings should undergo independent algorithmic impact analyses prior to implementation, thoroughly examining their accuracy, fairness, security, and potential effects on student rights. A regimen of periodic and trigger-based audits must be maintained to continuously monitor system performance, enabling the timely identification and rectification of biases or deviations.

Constructing a multifaceted governance ecosystem is critical, recognizing that oversight cannot rest solely on government or school authorities. A pluralistic co-governance model should prevail, comprising "government regulation, school accountability, industry self-discipline, parental participation, and societal oversight." Governments must promulgate baseline standards and policies; schools should establish internal ethics review committees; industry bodies ought to formulate technical ethics guidelines; parents must be endowed with rights to information and involvement; and third-party organizations should offer auditing and certification services. Through this diverse collaboration, a balanced dynamic is cultivated that both propels technological innovation and mitigates associated risks.

5. Conclusions

The empowerment of physical education evaluation through AI represents a profound paradigm shift. It propels evaluation from a static to a dynamic state, from a narrow to a holistic perspective, from mere certification to developmental guidance, and from a singular to a pluralistic approach, revealing immense potential to reshape the ecosystem of physical education. This study systematically elucidates the multifaceted significance of this transformation while candidly confronting the tangible challenges encountered in technological implementation, educational integration, and ethical security.

The research demonstrates that these challenges stem from the inherent tensions among technological rationality, educational philosophy, and ethical principles. Purely technological refinement cannot remedy the alienation within education; solely educational endeavor cannot guarantee data security; and delayed or inadequate governance risks stifling innovation. Hence, future progress must adhere to systemic thinking and a path of co-evolution. Technologically, the pursuit should be toward more trustworthy, interconnected, and contextually appropriate intelligence; educationally, it must unwaveringly center on holistic human development, establish a novel paradigm of human-AI collaboration, and safeguard the primacy of educators and the autonomy of learners; ethically and in governance, a forward-looking, rigorous, agile, and accountable framework of laws, regulations, and industry standards must be constructed to delineate safe boundaries for technological advancement.

References

[1] Li Yibing, Li Junfeng. Empowering School Physical Education Evaluation with Digital Intelligence Technology: Intrinsic Logic, Practical Dilemmas, and Quality Improvement Paths// School Physical Education Branch, Chinese Society of Sports Science. Proceedings of the 2024 National Academic Seminar for Young and Middle-aged School Physical Education Workers. School of Physical Education, Henan University, 2024: 153-154. DOI: 10.26914/c.cnkihy.2024.076946.

[2] Huang Chunchen, Lu Changfeng, Tian Youyi. Policy Evolution and Prospects of Artificial Intelligence Empowering Higher Education: A Three-Dimensional Analytical Framework Based on "Theme-Tool-Evaluation". Higher

Education Exploration, 2025, (01): 48-59.

[3] Ren Jinglun, Liang Shisuo. Artificial Intelligence + Physical Education Teaching: Key Technologies and Application Scenarios. Digital Teaching in Primary and Secondary Schools, 2024, (03): 10-15.

[4] Zhang Anyang. Reflections on Artificial Intelligence Empowering Reform of University Physical Education Teaching. Science & Technology of Stationery & Sporting Goods, 2025, (09): 172-174.

[5] Zheng Yonghe, Wang Yiyan, Yang Shuhao. The Value, Challenges, and Paths of Artificial Intelligence Empowering Education Evaluation. Open Education Research, 2024, 30(04): 4-10.

[6] Song Zhaoxiang, Si Linbo. The Logic, Boundaries, and Pathways of AI-empowered Educational Evaluation Reform in the New Era: Constructing a Transformative Path for Intelligent Educational Evaluation Ecosystems. Educational Science Research, 2024, (11): 59-65.

[7] Zhou Quan, Zhao Xinye. AI-Empowered Inclusive Physical Education: Value, Challenges, and Strategies. Sports Science and Technology Literature Bulletin, 2024, 32(10): 159-162+171. DOI: 10.19379/j.cnki.issn.1005-0256.2024.10.040.

[8] Yuan Shiwei, Mao Jie, Fu Zhihua, et al. Transformation of Physical Education Evaluation Empowered by Digital Technology: Connotation, Motivations, Problems and Strategies. Journal of Shenyang Sport University, 2024, 43(03): 31-38.

[9] Wu Ke, Zheng Qi. Advantages, Challenges and Approaches of Digitally Empowered School Physical Education Evaluation Work. Bulletin of Sport Science & Technology, 2024, 32(02): 214-217. DOI: 10.19379/j.cnki.issn.1005-0256.2024.02.056.

[10] Zhou Quan, Zhao Xinye. AI-Empowered Inclusive Physical Education: Value, Challenges, and Strategies. Bulletin of Sport Science & Technology, 2024, 32(10): 159-162+171.DOI:10.19379/j.cnki.issn. 1005-0256.2024.10.040.