Application Risks and Ethical Norm Construction of Artificial Intelligence in Educational Evaluation

Yangwenjun[12] Wuqianggang[12]

1 Mongolian National University， Ulaanbaatar city, Bayangol District, 11th Khoroo, National University of Mongolia，16060；

2 Jiangxi Institute of Fashion Technology， No. 103, Lihu Middle Avenue, Xiangtang Development Zone, Nanchang City, Jiangxi Province，330201；

Abstract： Artificial intelligence technology is driving the transformation of educational evaluation from empirical, outcome-based assessment to digital and process-oriented evaluation. However, risks arising from its application, such as technological overreach, algorithmic bias, and data security issues, pose potential threats to educational equity and humanistic values. Based on the policy direction of the "Overall Plan for Deepening Educational Evaluation Reform in the New Era," and integrating the technical logic of intelligent evaluation with the essence of education, this paper systematically analyzes four core risks of AI in educational evaluation. It proposes a three-dimensional ethical norm system of "technological optimization - institutional guarantee - subject empowerment" to provide theoretical support and practical pathways for the healthy development of intelligent educational evaluation.

Keywords： Ethical Risks; Norm Construction; Human-Machine Collaboration

DOI:10.69979/3041-0843.25.02.061

Introduction

Educational evaluation, serving as the "baton" for educational reform, directly impacts the quality of talent cultivation and the direction of educational development. The "Overall Plan for Deepening Educational Evaluation Reform in the New Era" issued by the Central Committee of the Communist Party of China and the State Council explicitly calls for "making full use of information technology to enhance the scientific, professional, and objective nature of educational evaluation," providing a policy basis for integrating AI into educational assessment. Through the collection and modeling of multi-source heterogeneous data, AI technology enables a comprehensive analysis of both the explicit states and implicit characteristics of evaluation subjects, promoting a shift towards more integrated and process-oriented educational evaluation.

However, while technological empowerment brings efficiency gains, it also triggers a series of ethical controversies. Existing research often focuses on the technical implementation of intelligent evaluation or discussions of single risks, lacking sufficient attention to the systematic nature of risks and the practicality of norms. Issues such as the tendency towards technocracy and algorithmic discrimination observed in some practices already reveal a deviation of technical logic from the essence of education. Therefore, systematically analyzing the application risks of AI in educational evaluation and constructing a suitable ethical norm system have become key propositions for balancing technological innovation and the educational foundation.

1 The Transformation Logic and Risk Dimensions of AI-Empowered Educational Evaluation

1.1 Core Transformation Logic of Intelligent Educational Evaluation

In the context of deepening digital education reform, artificial intelligence, with "data-driven + algorithmic modeling" at its core, has fundamentally reconstructed the implementation path of traditional educational evaluation. Its logical closed loop can be detailed into four stages: "indicator construction - data processing - result application - iterative

optimization," each deeply aligned with the essential needs of educational evaluation and technical characteristics.

In the indicator construction phase, intelligent evaluation is not a technological innovation detached from educational principles. It fundamentally follows the guidance of the "Overall Plan for Deepening Educational Evaluation Reform in the New Era," which aims to "overcome the sole focus on scores and academic advancement" and combines curriculum standards across different educational stages and subjects to build a multi-dimensional, hierarchical evaluation indicator system. For example, in Chinese language evaluation at the compulsory education stage, besides traditional quantitative indicators like vocabulary size and reading speed, qualitative indicators such as "text empathy ability" and "critical interpretation level" are added. These are transformed into analyzable feature variables through natural language processing technology, achieving an integrated "quantitative + qualitative" evaluation.

The data processing phase relies on multiple intelligent terminals to form a full-scenario collection network: Inside the classroom, smart interactive panels record students' response times, accuracy rates, and interaction frequencies; emotion recognition cameras analyze students' concentration levels and emotional fluctuations through facial micro-expressions. Outside the classroom, learning management systems track the completion paths, revision traces, and scope of extended reading in students' homework; some systems can even capture implicit data like writing pressure and pause rhythms through smart pens, reconstructing students' thinking processes. After cleaning, desensitization, and fusion, these multi-source heterogeneous data are used to build personalized evaluation models through machine learning algorithms – for instance, analyzing students' learning behavior preferences based on collaborative filtering algorithms, or assessing their comprehensive literacy development levels using deep learning models.

The result application phase breaks the limitation of traditional evaluation ending with a single assessment, upgrading the function from "diagnosis" to "improvement." AI learning analysis platforms generate personalized reports including "identification of knowledge weak points" and "suggestions for learning strategies." For example, regarding the problem of "low accuracy in geometry proof problems" in math learning, the system not only identifies the core reason as "insufficient reasoning in adding auxiliary lines" but also pushes similar variant exercises and micro-lecture resources. For teachers, the system outputs "suggestions for optimizing teaching segments," such as recommending situational teaching or group collaboration modes for cases where "classroom interaction participation rate for a certain knowledge point is low," forming a virtuous cycle of "promoting teaching and learning through evaluation."

In practical scenarios, intelligent evaluation has formed three mature application directions, covering the entire educational teaching chain: In the field of academic performance assessment, intelligent homework grading systems can instantly grade math calculation problems and English multiple-choice questions with over 98% accuracy, while also automatically scoring Chinese essays on dimensions like "structural completeness" and "language fluency," and flagging controversial answers for manual review. In learning state assessment, some schools have introduced "attention tracking systems" that generate real-time classroom attention heat maps, helping teachers dynamically adjust teaching pace to prevent learning effectiveness from being impacted by prolonged lapses in concentration. In teaching quality assessment, intelligent systems based on classroom video analysis can quantify teachers' questioning frequency, wait time, and interaction methods, combining this with student feedback data to form teaching quality reports, providing precise guidance for teachers' professional development.

However, if this technology-driven transformation loses the constraint of the "educating people" essence, it easily falls into the trap of "instrumental rationality supremacy." Some schools excessively pursue data completeness and evaluation efficiency, excluding key, hard-to-quantify aspects of the educational process – such as "teacher-student emotional communication" and "thought collisions in group collaboration" – from the evaluation scope. This even leads to alienated phenomena like "teaching for data" and "learning for indicators," contradicting the fundamental goal of educational evaluation to promote the holistic development of individuals.

1.2 Systematic Analysis of Four Core Application Risks

1.2.1 Risk of Technological Overreach: Obscuration and Alienation of the Educational Essence

Technological supremacism deifies intelligent evaluation as an "absolutely objective tool," leading to two major

misconceptions: First, the marginalization of the teacher's primary role. Some schools reduce the weight of teachers' experiential evaluation, allowing algorithmic scores (e.g., one middle school assigned AI teaching scores 60% weight in performance reviews) to dominate assessments, forcing teachers to adopt methods like "cramming + frequent questioning." Second, the oversimplification of evaluation content. To adapt to data collection, institutions reduce non-quantifiable activities like scientific inquiry and psychological counseling, while strengthening drill practices, deviating from the comprehensive education orientation.

1.2.2 Algorithmic Ethical Risks: Dual Impairment of Fairness and Validity

Risks stem from "data bias" and "algorithmic black boxes": Data bias leads to algorithmic discrimination. For instance, a system trained predominantly on data from urban key schools mistakenly labeled rural students' "dialect-influenced answers" as "logically confused," and through the "label reinforcement effect," restricted their access to resources. The algorithmic black box makes the decision-making process unexplainable. For example, if a student's essay is graded poorly by AI due to its novel viewpoint, and the technical provider evades responsibility citing "algorithmic neutrality," it undermines the trust foundation of evaluation.

1.2.3 Data Security Risks: Erosion of Privacy Rights Boundaries

Multi-dimensional sensitive data collection accompanies full-chain risks: During collection, schools bundle biometric data collection with system use, and institutions illegally collect facial and family information. During storage and transfer, data stored in plain text is vulnerable to attacks, and enterprises use educational data for commercial marketing. Furthermore, data like biometrics is irreplaceable; once leaked (e.g., facial information misused), it causes permanent damage to rights and interests.

1.2.4 Risk of Subject Alienation: Dissolution of Emotional Connection and Thinking Ability

Automated evaluation leads to "dehumanization": In teacher-student interaction, emotion monitoring systems reduce emotions to labels (e.g., "looking down = inattentive"), guiding teachers to respond mechanically, thus alienating teacher-student relationships. In student thinking, algorithms push "comfort zone" resources (e.g., only algebra exercises), creating information cocoons and weakening exploration and critical thinking abilities. Teachers, over-relying on AI suggestions, lose teaching autonomy and become "tool operators."

2 The Three-Dimensional System and Practical Pathways for Ethical Norm Construction

2.1 Technical Dimension: Building an Explainable, Unbiased Technical Foundation for Intelligent Evaluation

Technical optimization is the foundational support for ethical norms, requiring deep adaptation of algorithm design to educational needs. In the algorithm development phase, a collaborative mechanism involving "education experts - technical engineers" should be established to ensure the evaluation indicator system aligns with educational principles while being feasible for data representation. To address algorithmic bias, diverse training datasets incorporating educational samples from different regions and cultural backgrounds should be used, and discriminatory outputs reduced through adversarial training.

Promoting the development of explainable algorithms is key to solving the black box dilemma. Techniques like knowledge distillation and prototype networks should be used to transform the decision logic of complex algorithms into understandable rule explanations. Technical suppliers should be required to provide "algorithm specification documents," clarifying data sources, weight settings, and decision basis. An algorithm transparency certification system should be established, where third-party agencies test the fairness of intelligent evaluation systems, allowing only certified systems into educational settings.

At the data processing level, a full lifecycle security protection system should be built. Strictly adhere to the data minimization principle, collecting only information essential for evaluation. Employ encrypted storage and desensitization processing for sensitive data like biometrics. Establish a data hierarchical authorization mechanism, clearly defining the access boundaries for schools, teachers, and enterprises, and enabling full traceability and dynamic monitoring of data flow.

2.2 Institutional Dimension: Establishing a Multi-Stakeholder Collaborative Ethical Governance Framework

Institutional guarantee is crucial for norm implementation, requiring the construction of a multi-level governance system involving "government regulation - institutional autonomy - social supervision." At the national level, accelerate specialized legislation for intelligent educational evaluation, clarify data sovereignty ownership and privacy protection standards, and implement special licensing systems for biometric data collection. Drawing on the concept of "resilient + flexible" co-governance, establish a dynamically adjusted normative framework adaptable to technological iterations and practical developments.

Educational institutions need to establish internal ethical review mechanisms, forming review committees composed of teachers, parents, and ethics experts to oversee the introduction and use of intelligent evaluation systems throughout the entire process. Formulate norms for the use of evaluation results,clearly stipulating that algorithmic conclusions are for reference only, with teachers retaining final evaluation decision-making authority to prevent technological overreach. Establish complaint and appeal mechanisms, providing channels for teachers and students to seek redress for algorithmic bias and privacy violations.

Strengthen cross-subject collaborative supervision. Education administrative departments should take the lead in establishing a filing system for intelligent educational evaluation platforms, requiring enterprises to disclose data processing procedures and ethical commitments. Leverage industry associations to develop technical ethics conventions and practice guidelines, establish corporate credit archives, and impose market access restrictions on violators. Encourage public participation in supervision and enhance societal awareness of intelligent evaluation ethics through science popularization.

2.3 Subject Dimension: Cultivating Ethical Literacy and Capabilities for Human-Machine Collaboration

Subject empowerment is the implementation guarantee for ethical norms, requiring the enhancement of technological literacy and ethical awareness among educational participants. In teacher training systems, add courses on intelligent education ethics, covering algorithm principles, risk identification, and normative application. Integrate digital intelligence ethics capability into teacher professional development evaluation indicators. Through case-based teaching and simulated practices, cultivate teachers' judgment and decision-making abilities in human-machine collaboration scenarios, enabling them to effectively use technological tools while adhering to educational principles.

For the student group, implement digital citizenship education and critical thinking cultivation. Through curriculum teaching and practical activities, help students understand the working principles and potential limitations of intelligent evaluation, and foster awareness of data privacy protection. Encourage students to participate in discussions and formulation of evaluation standards, granting them the right to know and the right to object regarding the use of their personal data, avoiding excessive algorithmic intervention in learning autonomy.

Parents, as important participants, need to receive ethics literacy training through home-school cooperation channels, understanding the functional boundaries and risk points of intelligent evaluation systems, and playing a supervisory role in areas such as minor data authorization and interpretation of evaluation results. Form a subject collaboration pattern of "teacher guidance - student autonomy - parent supervision" to ensure technology application always serves the goal of educating people.

3 Conclusion

Artificial intelligence provides a new path of technological empowerment for educational evaluation reform, but its application risks are essentially products of the imbalance between technological logic and educational ethics. The core to avoiding risks and realizing value lies in adhering to the "education-oriented" principle, making the holistic development of individuals the fundamental direction of intelligent evaluation.

The "technological optimization - institutional guarantee - subject empowerment" three-dimensional ethical norm system constructed in this paper, through practical pathways of algorithm transparency, governance pluralism, and

comprehensive literacy, achieves a dynamic balance between technological innovation and ethical constraints. This system responds to practical challenges like technological overreach and algorithmic bias, while returning to the essential function of educational evaluation: promoting human development.

The future development of intelligent educational evaluation requires further deepening research on human-machine collaboration mechanisms, seeking a higher-level integration between technical efficiency and humanistic care. Through the rigid constraints of ethical norms and the flexible nourishment of value guidance, artificial intelligence can truly become an aid, not a hindrance, to educational evaluation reform, providing sustainable technical support for high-quality educational development.

References

[1] Lu Xingnan, Gao Xuewei. AI-Enabled Educational Evaluation Reform: Development Trends, Risk Examination, and Resolution Strategies[J]. China Higher Education Evaluation, 2023(3):54-59.
[2] Wu Chun. Research on the Application of Artificial Intelligence in Vocational Education Evaluation[J]. Microcomputer, 2024, 000(6):3.
[3] Fu Li. Research on the Application of Artificial Intelligence in Education[J]. Computer and Digital Engineering, 2012(012):040.
[4] Feng Rui; Sun Jiajing; Sun Faqin. Ethical Risks and Rational Choices in the Application of Artificial Intelligence in Education[J]. Journal of Distance Education, 2020(3).
[5] Yan Xiaopeng. Research on Ethical Risks and Prevention Strategies in the Educational Application of Artificial Intelligence Technology: Review and Prospects[C]// Proceedings of the "Integration of Industry and Education, School-Enterprise Cooperation" Education and Teaching Development Forum. 2024.