Available Online: 18 April 2026 DOI: 10.54254/3049-7248/2026.32883

Innovation and breakthroughs in basic education evaluation in the context of artificial intelligence

Caimin Zeng

East China University of Technology, Fuzhou, China

405263570@qq.com

Abstract. With the deep integration of artificial intelligence technologies and the digitalization of education, basic education evaluation is encountering a historic opportunity for paradigm transformation. Long-standing deficiencies in traditional evaluation systems—such as limited evaluative dimensions, the absence of process-oriented data, insufficient diagnostic precision, and weak guarantees of fairness—are being systematically addressed through the empowerment of intelligent technologies. Drawing on cutting-edge practices both domestically and internationally, this paper examines the innovative directions and breakthrough pathways for basic education evaluation in the age of artificial intelligence from four perspectives: evaluative philosophy, technological application, practical implementation, and ethical safeguards. It proposes an integrated developmental framework of "conceptual transformation—technological empowerment—practical implementation—ethical assurance", aiming to provide both theoretical guidance and practical reference for constructing a scientific, equitable, and efficient new ecosystem of basic education evaluation.

Keywords: artificial intelligence, basic education evaluation, evaluation innovation, digital transformation, core competencies

1. Introduction

Educational evaluation, as the "steering mechanism" of basic education reform, directly shapes the direction of talent cultivation and the improvement of educational quality. The *Overall Plan for Deepening the Reform of Educational Evaluation in the New Era* explicitly calls for "the use of modern information technologies such as artificial intelligence and big data to explore longitudinal, whole-process evaluation of students' learning across all grade levels, as well as horizontal, all-factor evaluation encompassing moral, intellectual, physical, aesthetic, and labor development" [1]. At present, China's basic education evaluation system continues to face a range of structural challenges. First, evaluation orientations tend to be utilitarian, with excessive emphasis on test scores and progression rates, while neglecting the cultivation of core competencies [2]. Second, evaluation methods remain dominated by summative paper-and-pencil testing, making it difficult to capture the dynamic developmental trajectories that unfold during the learning process [3]. Third, evaluation data are highly fragmented; the lack of interoperability across regions and platforms hinders data integration and the formation of scientifically grounded diagnoses [4]. Fourth, the evaluation process relies on a single entity and lacks a mechanism for diverse and collaborative participation [5].

The rise of artificial intelligence offers transformative tools to address these challenges. The application of generative AI, multimodal data collection, and machine learning algorithms is driving a shift in basic education evaluation—from experience-based judgment to data-driven analysis; from an exclusive focus on outcomes to a balanced emphasis on both process and results; and from single-dimensional assessment to multidimensional integration [6]. The *Six Pillars of Digital Transformation in Education: A Common Framework*, released by UNESCO, likewise emphasizes that digital technologies should serve as a core support for transforming educational evaluation, enabling greater precision, personalization, and fairness [4]. Against this backdrop, and drawing on the latest policy directives and academic research, this paper systematically examines the innovative dimensions and breakthrough pathways of basic education evaluation in the context of artificial intelligence, with the aim of providing both theoretical insight and practical support for advancing the high-quality development of basic education.

2. Conceptual innovation in basic education evaluation in the context of artificial intelligence

2.1. Transformation of evaluation objectives: from "knowledge-oriented" to "competency-oriented"

Traditional basic education evaluation has centered on the mastery of knowledge, relying on standardized testing to quantify students' academic performance. This approach has often led teaching into a vicious cycle of "teaching to the test" [2]. The application of artificial intelligence is driving a shift in evaluation objectives toward a "knowledge + competencies" orientation, with a focus on the holistic development of students' core competencies. This transformation does not negate the value of knowledge assessment; rather, it expands the evaluative focus to include dimensions that are less readily quantifiable, such as cognitive ability, innovative thinking, collaboration skills, and emotional attitudes [6].

Guided by the *Framework of Core Competencies for Chinese Students' Development,* AI-empowered evaluation systems construct a three-dimensional indicator model encompassing "cultural foundation–self-directed development–social participation", which is systematically aligned with the five domains of moral, intellectual, physical, aesthetic, and labor education [7]. For instance, natural language processing can be used to analyze students' classroom discourse and written texts in order to assess critical thinking and expressive ability; computer vision technologies can capture interaction patterns in group work to quantify collaborative competencies; and affective computing can monitor students' emotional states during the learning process to optimize personalized support strategies [8]. This competency-oriented approach represents a fundamental shift from "score-oriented education" to "holistic human development", aligning with the core objectives of building a strong education system [5].

2.2. Reconstruction of evaluation logic: from "outcome finality" to "integration of process and outcomes"

Traditional evaluation relies heavily on summative assessments conducted at a single point in time, making it difficult to reflect students' developmental trajectories and latent potential. Artificial intelligence breaks through these temporal and spatial constraints, enabling the construction of a "whole-process, whole-scenario" evaluation logic. Through multiple channels—including online learning platforms, smart classroom technologies, and intelligent terminals—data can be continuously collected across all stages of learning, such as pre-class preparation, in-class interaction, and post-class extension. These data encompass multiple dimensions, including learning duration, resource preferences, response trajectories, and error patterns. Once processed through algorithmic models, such data not only generate dynamic learning profiles but also provide a scientific basis for summative evaluation [6].

The organic integration of formative and summative evaluation lies at the core of this reconstructed logic. With machine learning algorithms, systems can predict students' summative performance based on process-oriented data, enabling early intervention in areas of weakness. At the same time, summative results feed back into the optimization of formative indicators, forming a closed-loop mechanism of "data collection–analytical diagnosis–feedback and improvement" [7]. For example, the SEED Comprehensive Student Quality Evaluation Platform integrates process-oriented data from three million students with summative assessment results to build an evaluation system that is both diagnostic and selective, enabling precise identification of developmental gaps while also supporting admission decisions [8].

2.3. Reshaping the evaluation landscape: from "single subject" to "multi-agent collaboration"

Traditional evaluations in basic education focus solely on teachers, making the results susceptible to subjective bias and unable to fully reflect students' development [5]. Artificial intelligence facilitates a transition toward a multi-agent collaborative model involving teachers, students, parents, technology systems, and society, thereby creating a more holistic and multi-perspective evaluation framework. Within this framework, each participant assumes distinct yet complementary roles: teachers provide targeted evaluation and feedback during instruction; students engage in self-assessment and peer assessment to develop metacognitive awareness; parents participate in process-oriented evaluation through intelligent platforms, strengthening home–school collaboration; AI systems undertake data collection, objective analysis, and preliminary diagnosis; and social institutions assess students' practical application abilities in real-world contexts [4].

The realization of such collaborative evaluation depends on the support of intelligent evaluation platforms. For instance, Chaoyang District in Beijing has developed an AI-empowered evaluation system that integrates four dimensions—teacher evaluation, student self-assessment, parental feedback, and platform-based diagnostics. Students' overall evaluation results are generated through weighted aggregation of multi-source data, ensuring both objectivity and broad participation [7]. This multi-agent collaborative model breaks the closed nature of traditional evaluation systems, advancing both the democratization and scientific rigor of educational assessment [3].

3. Technological breakthroughs in basic education evaluation in the context of artificial intelligence

3.1. Multimodal data acquisition: addressing the limitations of single-dimensional evaluation data

Traditional evaluation data have largely relied on quantitative scores derived from paper-and-pencil tests, making it difficult to capture students' non-cognitive competencies and practical abilities [5]. Artificial intelligence enables the comprehensive collection of multimodal data—including text, images, speech, and behavioral data—thereby providing a rich empirical foundation for holistic evaluation. Specifically, textual data are drawn from student assignments, essays, and classroom discussions, and can be analyzed using natural language processing to assess linguistic expression and depth of thinking. Image data, such as videos of experimental procedures, artistic works, and classroom behavior recordings, can be processed by computer vision techniques to extract key behavioral features. Speech data, collected from oral presentations and group discussions, can be evaluated through speech recognition technologies to assess fluency and communicative effectiveness. Behavioral data, including learning trajectories, resource selection, and problem-solving strategies, can be analyzed through learning analytics to uncover underlying learning patterns.

The integration of multimodal data enables a shift toward "de-surface-level" evaluation. For example, in the assessment of scientific experiments, intelligent systems can use cameras to capture students' procedural steps, adherence to operational norms, and collaborative interactions, and combine these with textual analysis of laboratory reports to provide a comprehensive evaluation of scientific inquiry competence [7]. In physical education, intelligent rope-skipping systems can collect data on endurance, balance, and coordination through sensors, transforming evaluation from mere "physical measurement" to "competency-based assessment" [7]. These technological advances convert previously intangible competencies into observable and analyzable data, thereby enhancing both the comprehensiveness and precision of evaluation [8].

3.2. Intelligent evaluation platforms: constructing a "data–diagnosis–feedback" closed loop

The deep application of artificial intelligence is driving evaluation platforms toward an integrated model of "data integration–intelligent analysis–precise feedback". In China, a number of representative intelligent evaluation platforms have emerged, such as the SEED Comprehensive Student Quality Evaluation and Development Platform and the National Smart Education Evaluation System for Primary and Secondary Schools. These platforms integrate core functions including data collection, algorithmic modeling, and results visualization, offering systematic solutions for basic education evaluation [8].

At the level of data integration, such platforms employ standardized interfaces to connect multiple data sources—such as student information systems, teaching management systems, and learning platforms—thus enabling cross-regional and cross-stage data interoperability [4]. For instance, Suzhou Industrial Park has implemented a mechanism of "indicator mapping, data integration, and application convergence" to integrate the SEED platform with existing regional information systems, forming a unified evaluation ecosystem [7]. At the level of intelligent analysis, platforms utilize machine learning algorithms and knowledge graphs to conduct deep mining of multimodal data, identifying students' strengths and developmental weaknesses [9]. For example, by constructing predictive learning models based on historical data, systems can forecast students' mastery of specific knowledge points, thereby informing personalized instruction. At the level of precise feedback, visualization technologies are used to present evaluation results in formats such as learning growth profiles and radar charts, enabling teachers, students, and parents to intuitively understand developmental status [6]. In Xi'an, for example, gamified assessment activities such as programming tasks have been integrated into routine teaching, with platforms providing real-time feedback on student progress, thereby enhancing engagement and motivation [7].

3.3. Optimization of algorithmic models: enhancing scientific rigor and fairness in evaluation

Algorithmic models constitute the core engine of AI-enabled educational evaluation, and their scientific validity directly determines the credibility of evaluation outcomes. At present, widely used algorithms in basic education evaluation include machine learning, deep learning, and natural language processing, all of which improve evaluation accuracy and fairness through large-scale data training [9]. For example, in the grading of subjective responses, AI systems can construct multidimensional evaluation frameworks to enable automated scoring of essays and open-ended questions. When combined with mechanisms such as averaging scores from multiple evaluators, these systems can significantly reduce subjective bias. In comprehensive quality evaluation, weighted algorithms can integrate multi-source and multi-dimensional data, preventing the overemphasis of any single indicator [7].

To address algorithmic bias, researchers have focused on optimizing both data sources and model design to ensure fairness [3]. On the one hand, expanding the coverage of data samples—by incorporating data from students across diverse regions and backgrounds—can mitigate discriminatory tendencies toward specific groups [9]. On the other hand, enhancing algorithmic transparency by disclosing core logic and indicator weightings to stakeholders allows for public scrutiny and accountability [7]. For instance, the learning analytics platform of Shandong Experimental High School has strengthened the credibility of its evaluation results by openly communicating its algorithmic principles, earning broad recognition from teachers, students, and parents [4].

4. Innovative practical pathways for basic education evaluation in the context of artificial intelligence

4.1. Exploration of region-specific "AI + evaluation" models

Given the significant disparities in the development of basic education across different regions of China, AI-empowered evaluation reform must be advanced in a context-sensitive manner. At present, pilot regions have developed diverse practical models, offering valuable experience for nationwide reform. In educationally advanced regions such as Beijing, Shanghai, and Suzhou, strong technological infrastructures have supported the establishment of evaluation models characterized by "comprehensive coverage and deep integration" [7]. For instance, Chaoyang District in Beijing has integrated SEED platform data with activities such as "Mathematics Festival" and "Science Festival" organized by innovation academies, enabling process-oriented evaluation to support talent identification [7]. In Shanghai, Luwan No. 1 Central Primary School in Huangpu District has developed an integrated "network–cloud–terminal" platform, deeply embedding intelligent devices and systems into teaching processes and promoting data-driven precision instruction [4].

In contrast, pilot regions in central and western China have focused on core needs, adopting a strategy of "targeted breakthroughs and phased advancement". For example, Guilin has linked data on higher-order thinking dimensions—such as problem-solving and innovation—with performance in science and technology competitions, providing multidimensional evidence for the early identification and cultivation of top innovative talents [7]. Heilongjiang Province has utilized dynamic data from 647,000 students to reveal group-level patterns and differences, facilitating a shift from experience-driven to data-driven evaluation [7]. These regional practices demonstrate that AI-enabled evaluation reform is not a one-size-fits-all process; rather, it requires alignment with local conditions within a unified framework to achieve contextually grounded development [4].

4.2. Large-scale application of value-added evaluation

Value-added evaluation, which focuses on students' developmental gains, aligns closely with the fundamental educational goal of fostering moral integrity and holistic development [2]. The application of artificial intelligence has enabled the large-scale implementation of value-added evaluation by tracking students' developmental data across different stages and quantifying the actual effects of educational interventions. Unlike traditional horizontal comparative evaluation, value-added evaluation emphasizes students' longitudinal growth, taking into account both their initial baseline and the magnitude of their progress. This approach provides a more objective reflection of educational quality and student potential [2].

In practice, value-added evaluation is often operationalized through regression models that control for external factors such as students' initial ability levels and family backgrounds, thereby isolating the contribution of educational interventions to developmental gains [3]. For instance, Jianlan Middle School in Hangzhou uses AI technologies to collect multidimensional data on students' learning interests, styles, and knowledge levels, constructing personalized learning pathways and evaluating intervention effectiveness through value-added metrics, which has effectively reduced students' academic burden [4]. Similarly, Yangyi No. 1 Primary School in Wenzhou applies the SOLO taxonomy (*Structure of the Observed Learning Outcome*) in mathematics competency assessment, tracking the progression of students' cognitive levels and achieving a deep integration of value-added evaluation with instructional improvement [4]. The adoption of value-added evaluation breaks the long-standing dominance of score-centric assessment, providing a robust scientific foundation for enhancing both educational equity and quality [2].

4.3. Innovation in interdisciplinary and practice-oriented evaluation

Artificial intelligence extends evaluation scenarios beyond the classroom into real-life contexts, thereby fostering innovation in interdisciplinary and practice-oriented assessment. Traditional evaluation has largely been confined to knowledge testing within single disciplines, making it difficult to assess students' capacity for integrated application [9]. AI-enabled evaluation addresses this limitation by creating authentic task-based scenarios that facilitate comprehensive assessment of interdisciplinary competencies. For example, in a "lunar exploration mission" evaluation scenario, students work collaboratively in a virtual simulation environment to design lunar probes, plan exploration routes, and respond to unexpected challenges. AI systems capture and analyze data in real time—such as team communication quality, operational sequences, and decision-making pathways—providing a comprehensive assessment of core competencies including innovative thinking and collaboration [7].

Innovation in practice-oriented evaluation is also reflected in the inclusion of previously underemphasized areas such as labor education and aesthetic education. Through intelligent terminals, data on students' practical activities—such as labor practice videos and artistic creations—can be collected and evaluated using a combination of expert judgment and AI analysis, enabling objective assessment of practical skills and aesthetic literacy [8]. For example, the "Seven-Color Sunshine Comprehensive Evaluation" system implemented at Tangxia Experimental Primary School in Rui'an constructs four evaluation dimensions—learning indicators, potential indicators, activity indicators, and self-directed indicators—to provide a holistic assessment of students' overall competencies [4]. These innovative practices realign evaluation with the fundamental purposes of education, promoting the all-round development of students [4].

5. Ethical safeguards and challenge responses in AI-empowered basic education evaluation

5.1. Core challenges: risks and dilemmas in technological application

While the application of artificial intelligence in basic education evaluation has generated significant breakthroughs, it also presents multiple challenges at the ethical, technical, and practical levels. First, there are risks related to data privacy and security. Sensitive data collected during evaluation processes—such as students' personal information and learning behaviors—are vulnerable to leakage and misuse [10]. In some cases, schools have collected data on students' facial expressions and attention levels without explicit informed consent, raising serious concerns about privacy protection [4]. Second, issues of algorithmic bias and fairness persist. If algorithmic models are trained on biased datasets, they may inadvertently amplify existing educational inequalities [3]. For instance, students in under-resourced or remote areas may be disadvantaged due to insufficient digital data, potentially leading to systematically lower evaluation outcomes [10]. Third, there is a risk of technological alienation. Overreliance on artificial intelligence may lead to a "technology-first" orientation in evaluation, overshadowing the humanistic dimensions of education [6]. In some contexts, teachers have delegated evaluative responsibilities entirely to intelligent systems, resulting in a lack of interpretive judgment and pedagogical intervention [7]. Fourth, the digital divide remains a critical concern. Significant disparities in technological infrastructure and teachers' digital literacy across regions and schools may further exacerbate inequalities in educational evaluation [10].

5.2. Response strategies: building a synergistic mechanism of "technological innovation–ethical regulation"

To address these challenges, it is essential to establish a governance framework that balances technological innovation with ethical regulation, ensuring the sustainable and responsible advancement of AI-enabled evaluation reform. First, data security governance systems must be strengthened by establishing comprehensive standards for data collection, storage, and usage [4]. Technical measures such as encrypted transmission, hierarchical access control, and data anonymization should be employed to safeguard student privacy. At the same time, the principle of data minimization must be strictly enforced, prohibiting the collection of irrelevant information without authorization [3]. Second, efforts should be made to enhance algorithmic fairness and transparency. This includes establishing algorithm auditing mechanisms to regularly assess fairness and mitigate bias [7]. Expanding the diversity of training datasets—by incorporating data from students of varied regions and backgrounds—can reduce structural bias, while disclosing core algorithmic logic and indicator weightings to stakeholders helps ensure the right to informed participation [9]. Third, improving teachers' digital literacy is essential. Through tiered and targeted professional development programs, teachers can strengthen both their capacity to apply AI technologies and their ethical judgment in evaluation practices [10]. For example, Zhongshan City in Guangdong Province has implemented regional teacher digital literacy training initiatives to enhance competencies in digital teaching and evaluation [4]. Similarly, the Miyun Branch of Beijing Yuying School has established a Young Teachers' Development Association to promote technical exchange and collaborative learning among educators [4]. Fourth, narrowing the digital divide requires increased investment in underdeveloped regions and disadvantaged schools. Measures such as school–enterprise partnerships and regional support programs can help improve technological infrastructure [10]. In addition, mechanisms for sharing high-quality evaluation resources should be established. Through national smart education platforms, exemplary evaluation tools and practices can be disseminated across regions, promoting more equitable access to advanced evaluation systems [4]. Through the coordinated advancement of these strategies, it is possible to ensure that artificial intelligence serves not only as a powerful technical enabler but also as a responsibly governed force in the transformation of basic education evaluation.

6. Conclusion and prospect

Artificial intelligence is reshaping the ecosystem of basic education evaluation. It is driving a transformation in evaluative philosophy from a "knowledge-oriented" paradigm to a "competency-oriented" one, advancing evaluation technologies from "single-dimensional approaches" to "multimodal integration", and guiding evaluation practices from "uniform models" toward "region-specific development". This transformation not only addresses the structural limitations of traditional evaluation systems but also aligns with the core requirements of talent cultivation in the broader agenda of building a strong education system. However, artificial intelligence is not a universal solution to all challenges in educational evaluation. Its application must remain grounded in the fundamental purposes of education, ensuring the organic integration of technological innovation and humanistic concern [6].

Looking ahead, the innovative development of basic education evaluation should focus on three key directions. First, it is necessary to develop education-specific large-scale models to enhance the precision and contextual relevance of evaluation, thereby reducing the biases associated with general-purpose models [6]. Second, the human–machine collaborative evaluation mechanism should be further strengthened, leveraging the advantages of artificial intelligence in data processing and objective analysis while preserving the central role of teachers in humanistic interpretation and value guidance [9]. Third, greater emphasis should be placed on the value-added application of evaluation outcomes, transforming evaluation data into a scientific basis for instructional improvement, resource allocation, and policy formulation, thus establishing a virtuous cycle of "evaluation–improvement–development" [7]. Through the coordinated advancement of technological innovation and institutional safeguards, artificial intelligence will provide sustained momentum for the evolution of basic education evaluation, contributing to the construction of a high-quality education system and laying a solid foundation for the development of a leading education nation.

References

[1] The State Council of the People's Republic of China. (2020, October 13). *Overall plan for deepening the reform of education evaluation in the new era [Z]*. https: //www.gov.cn/gongbao/content/2020/content_5554488.htm

[2] Li, K. Q. (2025). Reform of a high-quality English basic education evaluation system driven by artificial intelligence. *English Square,* (29), 129–132.

[3] Sun, B. B., & Chen, Z. N. (2025). Digital intelligence empowering basic education evaluation: Possibilities, challenges, and directions. *Reference for Basic Education,* (8), 15–23.

[4] Yin, S. F., Chen, B. E., Xu, B. B., et al. (2025). Practical dilemmas and breakthrough paths in the digital transformation of basic education evaluation: Insights from UNESCO's Six pillars of digital transformation in education: A common framework. *China Educational Technology,* (5), 85–94.

[5] Chang, S. L. (2025). Key focal points of basic education evaluation reform: Practical pathways based on the Outline for building a leading country in education (2024–2035). *Shanghai Journal of Educational Evaluation,* (5), 56–61.

[6] Liu, B. Q., Nie, X. L., Wang, S. J., et al. (2024). Generative artificial intelligence and the reshaping of future education: Technical frameworks, capability characteristics, and application trends. *e-Education Research,* (1), 13–20.

[7] Wu, Y., Li, L. Y., & Zheng, Q. H. (2025). Theoretical reconstruction and practical pathways of artificial intelligence empowering students' comprehensive quality evaluation. C*hina Educational Informatization,* (10), 111–118.

[8] Zheng, Q. H. (2025). Artificial intelligence empowering reform and innovation in education evaluation. *Journal of Wuhan University (Natural Sciences Edition)*, (5), 589–590.

[9] Li, Y. F. (2025). International experience in AI-empowered reform of higher education evaluation. *e-Education Research,* (2), 32–40.

[10] Zeng, Q. L. (2025). Intelligent technology empowering higher education curriculum management: Challenges, pathways, and countermeasures. *China-Arab Science and Technology Forum,* (10), 116–120.