

AI-Driven Education Evaluation Reform: From Quantitative Assessment to Qualitative Empowerment

Meng Wang

Xi'an Technological University, Xi'an, Shaanxi, China

2663526250@qq.com

Abstract. As the "navigator" and "calibrator" of educational activities, education evaluation directly determines the direction of educational goals and the quality of talent training. For a long time, education evaluation in China has been dominated by quantitative assessment. Although it has achieved standardization and efficiency in evaluation, it is difficult to take into account the consideration of students' qualitative literacy such as higher-order thinking, essential character, and emotional attitude, falling into the dilemma of "score-only theory". The iterative upgrading of artificial intelligence (AI) technology has provided a historic opportunity to solve this dilemma, promoting the transformation of education evaluation from a quantitative orientation of "digital ruler" to a qualitative empowerment of "situational insight". Based on the practical needs of education evaluation reform, this paper analyzes the limitations of traditional quantitative assessment and the advantages of AI-enabled qualitative evaluation, constructs an AI-driven integrated "quantitative + qualitative" education evaluation system, verifies the implementation path with practical cases and specific charts, and discusses the challenges and solutions in the reform process, providing theoretical reference and practical guidance for education evaluation reform in the new era.

Keywords: AI Technology; Education Evaluation Reform; Quantitative Assessment; Qualitative Empowerment; Core Literacy

1. Introduction

The convening of the 2024 National Education Conference and the promulgation and implementation of the "Outline for Building an Education Power (2024—2035)" at the beginning of 2025 have provided a programmatic guideline for deepening the comprehensive reform in the field of education. Among them, "establishing a scientific education evaluation orientation" is placed at the core, highlighting that building a new evaluation system supporting "comprehensive improvement of independent talent training quality and continuous emergence of top-notch innovative talents" has become a strategic fulcrum for building an education power [1]. The core proposition of education evaluation reform has always centered on the two issues of "what to evaluate" and "how to evaluate". However, the traditional education evaluation model takes quantitative assessment as the core, simplifying the complex educational process into measurable and calculable digital indicators, which is difficult to adapt to the educational goals of "simultaneous development of morality, intelligence, physical education, aesthetic education and labor education" and core literacy training in the new era.

With the in-depth penetration of AI technologies (such as machine learning, computer vision, speech emotion analysis, and large language models) in the field of education, education evaluation is experiencing a paradigm shift from empirical judgment to data-driven, and from static description to dynamic process tracking [1]. AI technology can break through the limitations of traditional evaluation methods, realize multi-dimensional and whole-process data collection of students' cognitive processes, behavioral performance and emotional attitudes, provide accurate data support for qualitative evaluation, and promote the upgrading of education evaluation from "single quantitative" to "integration of quantitative and qualitative". Based on this, this paper focuses on the

reform path of AI-driven education evaluation, focuses on how to rely on AI technology to realize the transformation from quantitative assessment to qualitative empowerment, solve the pain points of traditional education evaluation, and construct a scientific, comprehensive and developmental new paradigm of education evaluation.

2. Dilemmas of Traditional Quantitative Assessment and Highlighted Value of Qualitative Evaluation

2.1 Core Limitations of Traditional Quantitative Assessment

Traditional education evaluation mainly adopts quantitative assessment, which focuses on standardized measurement of educational effects through quantifiable indicators such as exam scores, attendance rates, and homework completion rates. Its advantages lie in strong objectivity, simple operation, and strong comparability, providing effective support for education quality monitoring in a specific historical period. However, with the deepening of education reform, this single quantitative assessment model has gradually exposed many limitations, which can be specifically summarized into the following three points:

First, the evaluation dimension is narrow, ignoring qualitative literacy. Quantitative assessment focuses on students' mastery of knowledge and emphasizes the measurement of "knowledge and skills", but it is difficult to cover students' qualitative literacy such as critical thinking, innovation ability, cooperation awareness, emotional attitude, and values [4]. For example, in traditional evaluation, students' composition evaluation is mostly measured by quantitative indicators such as word count, typos, and grammatical norms, ignoring the ideological depth, logical consistency, and emotional expression of the article, resulting in evaluation results that are difficult to reflect students' real writing ability and thinking level.

Second, the evaluation process is static, lacking dynamic tracking. Quantitative assessment is mostly carried out in static ways such as final exams and phased assessments, focusing on the evaluation of educational results, and it is difficult to capture students' growth changes, progress trajectories, and potential abilities in the learning process [4]. For example, if a student's exam score is unsatisfactory, quantitative assessment can only present the result of "low score", but cannot analyze the underlying reasons (such as knowledge gaps, improper learning methods, psychological pressure, etc.), making it difficult to provide accurate basis for personalized teaching intervention.

Third, the evaluation subject is single, falling into "digital tyranny". Traditional quantitative assessment is mostly led by teachers, with scores as the core evaluation standard. Students are in a passive position of accepting evaluation, lacking opportunities for self-expression and self-reflection [4]. At the same time, excessive reliance on quantitative indicators is likely to lead to "score-only theory". Teachers ignore the cultivation of students' all-round development in pursuit of scores, and students fall into the dilemma of "examination-oriented education" in pursuit of scores, distorting the essential value of education.

2.2 Core Value of Qualitative Evaluation

Qualitative evaluation takes "people-oriented" as the core. Through various methods such as observation, interview, work analysis, and situational test, it deeply understands students' learning process, thinking mode, emotional attitude, and values, and focuses on the comprehensive consideration of students' personality development and comprehensive literacy. Its core value is reflected in the following two aspects:

On the one hand, qualitative evaluation can make up for the dimensional defects of quantitative assessment and realize the comprehensive evaluation of students' core literacy. Qualitative evaluation not only pays attention to students' mastery of knowledge, but also pays more attention to students' qualitative literacy that is difficult to quantify, such as thinking ability, innovation awareness, and cooperation ability, which can more truly and comprehensively reflect students' comprehensive development level [4]. For example, by analyzing students' inquiry learning reports,

creative works, and classroom speeches, we can capture students' innovative thinking and practical ability, and realize the accurate portrayal of students' core literacy.

On the other hand, qualitative evaluation focuses on process tracking and can provide support for personalized education. Qualitative evaluation runs through the entire learning process. By continuously observing and recording students' learning behaviors and growth changes, it can accurately identify students' learning needs, advantages, and deficiencies, provide a basis for teachers to formulate personalized teaching plans and implement targeted interventions, and truly realize "evaluation promoting learning" and "evaluation promoting teaching" [4].

2.3 Complementary Logic between Quantitative and Qualitative Evaluation

Quantitative assessment and qualitative evaluation are not opposites, but "two sides of a coin", and they have a clear complementary logic [4]. Quantitative assessment provides data support for qualitative evaluation, avoiding it from falling into vague subjective judgment; qualitative evaluation endows meaning to quantitative assessment and explains the educational logic behind the data. For example, in the evaluation of students' extracurricular reading, we not only count the per capita reading volume (quantitative indicator), but also analyze reading notes and carry out reading sharing sessions (qualitative analysis) to find that some students have a large reading volume but prefer fast-food reading materials, thereby adjusting reading guidance strategies and realizing the scientificity and comprehensiveness of evaluation.

Table 1 Comparative Analysis of Quantitative Assessment and Qualitative Evaluation

Evaluation Type	Core Characteristics	Evaluation Methods	Advantages	Limitations
Quantitative Assessment	Standardized, Objective, Quantifiable	Examinations, Tests, Data Statistics	Simple Operation, Strong Comparability, High Objectivity	Narrow Dimension, Ignoring Qualitative Literacy, Static and One-sided
Qualitative Evaluation	Personalized, Contextualized, Process-oriented	Observation, Interview, Work Analysis, Situational Test	Comprehensively Reflect Core Literacy, Focus on Process Tracking	Strong Subjectivity, Complex Operation, Lack of Unified Standards

3. Feasibility and Advantages of AI-Driven Education Evaluation Transformation from Quantitative to Qualitative Empowerment

3.1 Feasibility Analysis

The core advantages of the iterative upgrading of AI technology lie in the comprehensiveness of data collection, the accuracy of data analysis, and the intelligence of the evaluation process, which are highly consistent with the needs of qualitative evaluation, providing a solid technical support for the transformation of education evaluation from quantitative to qualitative empowerment. Its feasibility is mainly reflected in two aspects:

First, AI technology can realize comprehensive collection of multi-modal data, solving the problem of qualitative evaluation data acquisition. Qualitative evaluation requires a large amount of process data (such as classroom behavior, speech content, work details, emotional changes, etc.). In the traditional way, it is difficult for teachers to comprehensively and real-time record these data, leading to the difficulty in implementing qualitative evaluation [2]. AI technology, through tools such as cameras, audio equipment, and learning terminals, can real-time collect multi-modal data such as students' classroom interaction frequency, speech duration, expression changes, and operation processes, automatically store and sort them, providing sufficient data support for qualitative evaluation [3]. For example, in the classroom demonstration activity, Wenzhou Xiedu

No.1 Primary School used AI classroom analysis software to real-time capture teacher-student interaction data and students' speech duration, generating a classroom participation heat map, which provided accurate data reference for qualitative evaluation [3].

Second, AI technology can realize intelligent data analysis, improving the objectivity and accuracy of qualitative evaluation. The core pain point of qualitative evaluation is strong subjectivity and lack of unified standards. AI technology, through algorithms such as machine learning and natural language processing, can intelligently analyze the collected multi-modal data, convert vague qualitative characteristics into quantifiable analysis results, and reduce the interference of human subjective factors [1]. For example, analyzing students' composition content through natural language processing technology can accurately identify the ideological depth, logical consistency, and emotional expression of the article; analyzing students' classroom expressions and body movements through computer vision technology can judge students' learning status and emotional attitude [2].

3.2 Core Advantages of AI-Driven Qualitative Empowerment

Compared with traditional qualitative evaluation methods, AI-driven qualitative empowerment has obvious advantages, which can effectively solve the pain points of traditional qualitative evaluation and promote the improvement of education evaluation quality and efficiency. The specific advantages are as follows:

First, intelligent data processing improves evaluation efficiency. Traditional qualitative evaluation requires teachers to spend a lot of time sorting out and analyzing process data, which is inefficient. AI technology can automatically complete data collection, classification, analysis, and sorting, freeing teachers from tedious transactional work, allowing teachers to devote more energy to the application of evaluation results and personalized teaching intervention [2]. For example, the AI teaching evaluation system can automatically generate teachers' classroom teaching behavior analysis reports without teachers manually counting data such as teacher-student interaction time and question types, greatly improving evaluation efficiency [2].

Second, precise evaluation diagnosis realizes personalized empowerment. Through continuous tracking and analysis of students' learning process data, AI technology can accurately identify students' learning advantages, deficiencies, and personalized needs, build personalized evaluation files for each student, and realize "one evaluation per person" [2]. For example, the SEED platform developed by Beijing Normal University, by integrating machine learning, computer vision and other technologies, realizes the accurate diagnosis and objective evaluation of students' cross-disciplinary abilities, character, and values, providing support for personalized education [1].

Third, process-oriented dynamic tracking reflects developmental evaluation. AI technology can real-time capture students' learning process data, dynamically update evaluation results, clearly present students' growth trajectories and progress changes, break the limitation of "one exam determines a lifetime" in traditional evaluation, and reflect the evaluation concept of "treating students from a developmental perspective" [1]. For example, the AI system can establish an exclusive professional development electronic file for each teacher, continuously record teaching videos, evaluation reports, improvement results and other data in different periods, and clearly present the teacher's growth trajectory [2].

Fourth, diversified evaluation subjects realize collaborative empowerment. AI technology can build a diversified evaluation platform, integrate the opinions of multiple evaluation subjects such as teachers, students, and parents, and realize the comprehensiveness and objectivity of evaluation [2]. For example, students can submit self-reflection reports through the AI platform, parents can feedback students' family learning performance through the platform, and teachers form comprehensive and objective evaluation results by combining AI analysis data and multi-party feedback, realizing "multi-party collaboration and comprehensive evaluation".

4. Practical Path of AI-Driven Education Evaluation Transformation from Quantitative to Qualitative Empowerment

Based on the advantages of AI technology and the needs of education evaluation reform, this paper constructs an AI-driven integrated "quantitative + qualitative" education evaluation system, clarifies the practical path of education evaluation transformation from quantitative assessment to qualitative empowerment from four links: data collection, data analysis, evaluation implementation, and result application, and ensures the operability and practicality of the path by combining specific charts and practical cases.

4.1 Multi-Modal Data Collection: Building a Comprehensive Evaluation Data Base

Data collection is the foundation of AI-driven qualitative empowerment. It is necessary to break through the limitations of traditional quantitative data, build a multi-modal data collection system of "quantitative data + qualitative data", and ensure the comprehensiveness, real-time performance, and accuracy of data. The specific collection content and methods are as follows:

First, quantitative data collection retains the advantages of traditional quantitative assessment, focuses on students' knowledge mastery and learning foundation, mainly including exam scores, homework completion rates, attendance rates, knowledge point mastery, etc., which are automatically collected through learning terminals, educational administration systems and other tools to ensure the objectivity and standardization of data [4].

Second, qualitative data collection, relying on AI technology to achieve comprehensive coverage, focuses on students' core literacy and learning process, mainly including classroom behavior data (speech content, interaction frequency, operation process), work data (compositions, creative works, inquiry reports), emotional data (expression changes, intonation), reflection data (self-summary, learning experience), etc. [2]. For example, collect students' classroom participation, question quality and other data through AI classroom analysis software, collect students' work content and reflection reports through natural language processing technology, and collect students' emotional changes through computer vision technology [3].

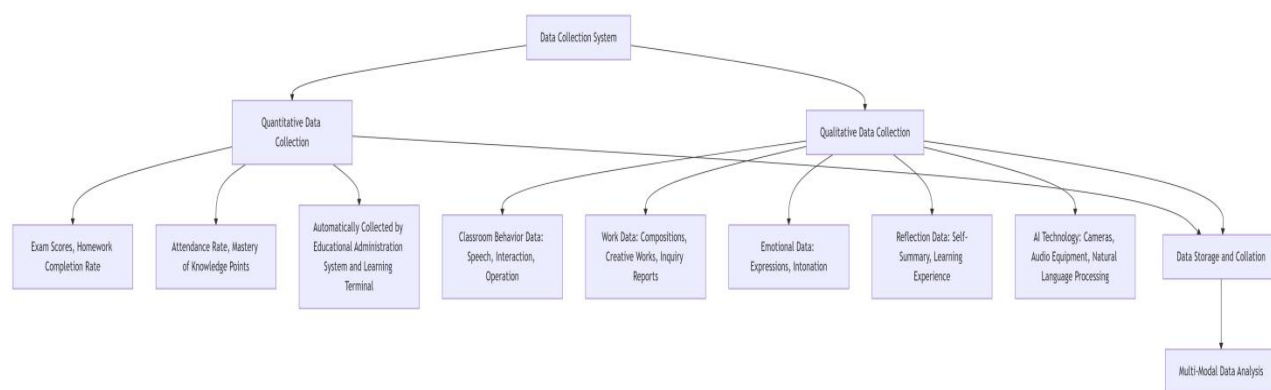


Figure 1. Framework of AI-Driven Multi-Modal Data Collection System

Case Evidence: In the "Intelligent Empowerment, AI Diagnosis" classroom demonstration activity, Wenzhou Xiedu No.1 Primary School collected multi-modal data of 12 courses of different subjects and grades through AI classroom analysis software, including teacher-student interaction frequency, students' speech duration, attention curve, homework completion quality, etc., providing comprehensive data support for qualitative evaluation and realizing accurate portrayal of students' classroom performance [3].

4.2 Intelligent Data Analysis: Realizing In-depth Integration of Quantitative and Qualitative Data

Data analysis is the core of AI-driven qualitative empowerment. It is necessary to rely on AI algorithms to realize in-depth integration of quantitative data and qualitative data, convert vague qualitative characteristics into analyzable and applicable evaluation results, and improve the accuracy and objectivity of evaluation. The specific implementation methods are as follows:

First, quantitative data analysis. Through statistical analysis, comparative analysis and other methods, explore the laws behind quantitative data. For example, by analyzing students' exam scores, identify students' knowledge gaps and weak links; by comparing score changes in different stages, track students' learning progress trajectory [4].

Second, qualitative data analysis, which is intelligently processed by AI algorithms. For example, analyze students' compositions and reflection reports through natural language processing technology, extract core viewpoints, logical structures and emotional tendencies, and quantitatively evaluate students' thinking ability and emotional attitude; analyze students' classroom expressions and body movements through computer vision technology to judge students' learning status (focused, tired, distracted); analyze students' work data through machine learning algorithms to evaluate students' innovation ability and practical ability [1].

Finally, integrated analysis. Combine the results of quantitative data analysis and qualitative data analysis to form a comprehensive and objective evaluation conclusion. For example, a student has excellent exam scores (quantitative data), but through AI analysis of his classroom speeches and inquiry reports, it is found that his innovation ability and cooperation ability are insufficient (qualitative data). The evaluation conclusion needs to take into account both his knowledge mastery and core literacy, providing an accurate basis for personalized teaching intervention [4].

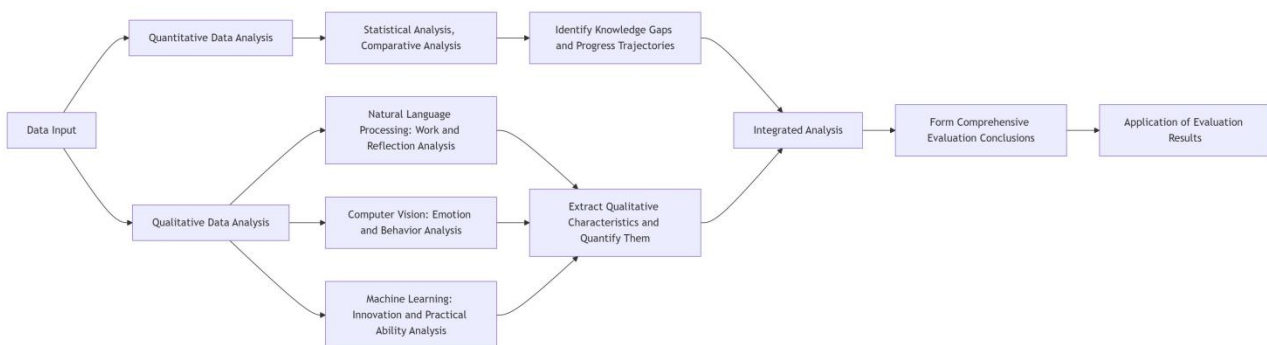


Figure 2. Flow Chart of AI-Driven Integrated Analysis of Quantitative and Qualitative Data

Case Evidence: The SEED platform developed by Beijing Normal University integrates machine learning, computer vision, speech emotion analysis, and large language models enhanced by expert knowledge, realizing accurate mapping and interpretable evaluation from underlying behavioral data to high-level literacy dimensions. For example, in the "Scientific Exploration Journey" task, the experimental operation process of students is analyzed through behavior recognition technology to evaluate their scientific inquiry ability; in the "Little Artist" task, the large model embedded with expert knowledge is used to conduct multi-dimensional evaluation of art works, effectively overcoming the deviation of traditional subjective evaluation [1].

4.3 Diversified Evaluation Implementation: Building an "AI + Multi-Subject" Evaluation Model

The evaluation implementation link needs to break the limitation of traditional single teacher evaluation, rely on the AI platform to build an "AI + teacher + student + parent" diversified evaluation model, and realize the comprehensiveness and objectivity of qualitative evaluation. The specific implementation process is as follows:

First, AI intelligent evaluation. Rely on AI algorithms to analyze multi-modal data and generate preliminary qualitative evaluation results, including evaluation of students' thinking ability,

innovation ability, emotional attitude and other dimensions, providing reference for subsequent evaluation [2].

Second, teacher evaluation. Combining the evaluation results generated by AI with their own teaching observation and experience, teachers supplement the evaluation of students' qualitative literacy, correct the deviations in AI evaluation, and ensure the objectivity and comprehensiveness of evaluation results [2]. For example, if AI identifies that a student's classroom attention is declining, the teacher judges that the reason is the high difficulty of the examples based on classroom memory, and then supplements the evaluation opinions to provide a basis for teaching improvement [2].

Third, student self-evaluation. Through the AI platform, students check their own learning process data and AI evaluation results, conduct self-reflection, summarize their own advantages and disadvantages, and submit self-evaluation reports, cultivating students' self-reflection ability and autonomous learning awareness [2].

Fourth, parent evaluation. Parents understand students' learning situation and evaluation results through the AI platform, submit parent evaluation opinions combined with students' family learning performance, realize home-school collaborative evaluation, and form an educational joint force [2].

Table 2 Implementation Content and Subjects of AI-Driven Diversified Evaluation Model

Evaluation Subject	Evaluation Content	Evaluation Methods	Core Role
AI	Thinking Ability, Innovation Ability, Emotional Attitude, Learning Status	Multi-Modal Data Analysis, Algorithm Modeling	Provide Accurate Data Support and Improve Evaluation Efficiency
Teacher	Comprehensive Literacy, Learning Attitude, Progress and Changes	Classroom Observation, Empirical Judgment, Supplementary Evaluation	Correct AI Deviations and Ensure Evaluation Objectivity
Student	Self-Cognition, Advantages and Disadvantages, Learning Reflection	Self-Evaluation Report, Process Reflection	Cultivate Self-Reflection and Autonomous Learning Ability
Parent	Family Learning Performance, Living Habits, Emotional Quality	Submission of Evaluation Opinions, Home-School Communication	Realize Home-School Collaboration and Form Educational Joint Force

4.4 Application of Evaluation Results: Realizing "Evaluation Promoting Learning and Teaching"

The application of evaluation results is the ultimate goal of education evaluation reform. It is necessary to break the dilemma of "disconnection between evaluation and teaching", deeply integrate the AI-driven qualitative evaluation results with teaching practice, and realize "evaluation promoting learning and evaluation promoting teaching". The specific application paths are as follows:

First, for students: personalized learning guidance. According to the evaluation results generated by AI, formulate personalized learning plans for each student, and provide targeted learning resources and guidance strategies for students' weak links and qualitative literacy shortcomings [2]. For example, for students with insufficient innovation ability, recommend inquiry learning tasks and

creative practice activities; for students with negative emotional attitudes, strengthen psychological counseling and encouragement to help students establish learning confidence. At the same time, push evaluation results and growth suggestions to students through the AI platform, allowing students to clearly understand their own progress and deficiencies, and stimulate their learning motivation [2].

Second, for teachers: teaching optimization and professional development. AI-driven qualitative evaluation results can provide teachers with accurate teaching feedback, help teachers identify problems and deficiencies in the teaching process, and optimize teaching design and teaching methods [2]. For example, by analyzing students' classroom behavior data, teachers can find problems such as insufficient interaction and single questioning methods in their own teaching, and then adjust teaching strategies to improve teaching quality. At the same time, the AI system can track and analyze teachers' teaching data for a long time, identify teachers' professional development shortcomings, and push personalized training resources to promote teachers' professional growth [2].

Third, for education managers: educational decision-making and quality monitoring. AI-driven education evaluation can provide education managers with comprehensive and accurate education quality data, help managers grasp the current situation of regional education development, and formulate scientific educational policies and decisions [1]. For example, by analyzing the core literacy evaluation data of students in the region, managers can identify the weak links in regional education development, optimize the allocation of educational resources, and promote the high-quality development of regional education.

Case Evidence: Since 2022, the SEED platform has been applied on a large scale in more than 8,000 primary and secondary schools in 44 experimental areas of 29 provinces (autonomous regions and municipalities directly under the Central Government) across the country, serving more than 3 million students in grades 3 to 11 for three consecutive years. Through the in-depth empowerment of AI technology, it has successfully explored and practiced the "diagnosis" and "identification" functions of comprehensive quality evaluation, providing strong support for teachers' teaching optimization, students' personalized development and education managers' decision-making [1].

5. Challenges and Solutions in AI-Driven Education Evaluation Reform

5.1 Main Challenges Faced

Although AI technology provides strong support for the transformation of education evaluation from quantitative to qualitative empowerment, it still faces many challenges in the practice process, which can be specifically summarized into the following four points:

First, technical level: limitations of AI algorithms and data security issues. At present, AI algorithms still have certain limitations, and it is difficult to fully and accurately interpret complex qualitative characteristics such as students' emotional attitudes and values, which may lead to evaluation deviations [2]. At the same time, in the process of multi-modal data collection, it involves students' personal information, learning data and other private content. If the data storage and management are improper, data leakage may occur, infringing on students' privacy [2].

Second, practical level: insufficient AI application ability of teachers and formalization of evaluation. Some teachers lack AI technology application ability, making it difficult to skillfully operate AI evaluation tools and effectively use the evaluation results generated by AI to optimize teaching practice [2]. At the same time, when some schools promote AI-driven qualitative evaluation, there is a problem of "emphasizing form over effect". They only simply introduce AI tools, but do not deeply integrate the evaluation results with teaching practice, leading to the formalization of qualitative evaluation [4].

Third, conceptual level: the constraint of traditional evaluation concepts. For a long time, the evaluation concept of "score-only theory" has been deeply rooted in people's hearts. Some teachers and parents still take exam scores as the core standard for evaluating students, lack recognition and

attention to AI-driven qualitative evaluation, and are difficult to actively cooperate with the promotion of evaluation reform [4].

Fourth, ethical level: lack of fairness and humanistic care in AI evaluation. The training data of AI algorithms may have deviations, leading to fairness issues in evaluation results. For example, discriminatory evaluations may be made against students from different family backgrounds and different learning foundations [2]. At the same time, AI evaluation over-reliance on data and algorithms may ignore students' individual differences and humanistic care, leading to evaluation becoming cold and mechanical [2].

5.2 Solutions

In response to the above challenges, combined with the needs of education evaluation reform, the following solutions are proposed to promote the orderly progress of AI-driven education evaluation reform:

First, technological optimization: improve AI algorithms and strengthen data security guarantees. On the one hand, increase investment in AI technology research and development, optimize AI algorithm models, combine educational and teaching rules and students' development characteristics, improve AI's ability to interpret qualitative characteristics, and reduce evaluation deviations [1]. For example, in AI algorithm training, integrate the experience and suggestions of educational experts to enhance the educational adaptability of the algorithm. On the other hand, establish and improve data security management systems, use encryption technology to store and protect students' personal data, clarify data usage rights, prevent data leakage, and ensure students' privacy and security [2].

Second, teacher empowerment: strengthen AI application training and improve teachers' professional abilities. Education departments and schools should carry out targeted AI technology application training to help teachers master the operation methods of AI evaluation tools and improve teachers' ability to analyze and apply AI evaluation results [2]. At the same time, establish an incentive mechanism for teachers' AI application, encourage teachers to actively explore the integration path of AI and education evaluation, and deeply integrate evaluation results with teaching practice to avoid the formalization of evaluation [3].

Third, concept renewal: strengthen publicity and guidance, and establish a scientific evaluation concept. Through various methods such as parent meetings, teacher training, and education publicity, popularize the concept of AI-driven qualitative evaluation to teachers and parents, break the constraint of "score-only theory", let teachers and parents realize the important significance of qualitative evaluation for students' all-round development, and actively cooperate with the promotion of evaluation reform [4].

Fourth, ethical norms: establish AI evaluation ethical guidelines to ensure evaluation fairness and humanistic care. Formulate ethical guidelines for AI education evaluation, standardize the design and application of AI algorithms, and avoid fairness issues caused by algorithm deviations [2]. At the same time, emphasize the "AI + manual" dual-track evaluation mechanism. On the basis of AI evaluation, give full play to teachers' subjective initiative, pay attention to students' individual differences and humanistic care, and avoid the mechanicality and coldness of evaluation [2]. For example, the "Famous Teacher Growth Project Team" of the Chinese Academy of Educational Sciences has implemented a mixed evaluation model of "AI + experts + researchers + teachers + students", which not only improves evaluation efficiency, but also ensures the element of humanistic care [2].

6. Conclusion and Outlook

The iterative upgrading of AI technology has provided new opportunities for education evaluation reform, promoting the transformation of education evaluation from quantitative assessment to qualitative empowerment, solving the dilemma of "score-only theory" in traditional education evaluation, realizing comprehensive and accurate evaluation of students' core literacy, and

providing strong support for "evaluation promoting learning and evaluation promoting teaching". By analyzing the limitations of traditional quantitative assessment and the value of qualitative evaluation, this paper clarifies the feasibility and advantages of AI-driven education evaluation transformation from quantitative to qualitative empowerment, constructs a complete practical path of "data collection - data analysis - evaluation implementation - result application", verifies the operability and practicality of the path with specific charts and practical cases, and puts forward corresponding solutions for the challenges faced in the reform process.

Looking forward to the future, with the in-depth integration of cutting-edge technologies such as multi-modal perception, emotion computing, and neuroscience, AI will more accurately and invisibly capture implicit thinking processes, collaboration states, and innovation potential, realizing scientific measurement of higher-order thinking and implicit character [1]. The AI-driven education evaluation reform will be further deepened, realizing the in-depth integration of quantitative and qualitative, and constructing a more scientific, comprehensive and developmental new paradigm of education evaluation. At the same time, it is necessary to strengthen the in-depth integration of AI technology with educational and teaching rules, focus on humanistic care and fairness, so that AI can truly serve the all-round development of students and the high-quality development of education, and provide strong support for building an education power.

References

- [1] Zheng, Q. (2025). AI empowers education evaluation reform and innovation. *Journal of Wuhan University (Natural Science Edition)*, 71(5), 589-590.
- [2] Qing, S. (2026). How to use AI evaluation to improve teaching. *Chinese Academy of Educational Sciences*, (2), 1-5.
- [3] Lucheng District Institute of Educational Research, Wenzhou City. (2025). Intelligent empowerment, AI diagnosis - Classroom demonstration activity based on the application of AI classroom analysis software.
- [4] Anonymous. (2025). *Methods for combining quantitative indicators and qualitative analysis in education evaluation*. Beijing: Educational Science Press.
- [5] Zhong, Q. (2023). Education evaluation reform oriented by core literacy. *Educational Research*, 44(1), 10-20.
- [6] Wu, G. (2024). Ethical dilemmas and path reconstruction of AI-enabled education evaluation. *Global Education Outlook*, 53(3), 3-15.
- [7] Li, M., & Shi, J. (2024). From quantitative to qualitative: The transformation logic and practical path of AI-driven education evaluation. *Educational Research and Experiment*, (2), 78-83.
- [8] Ministry of Education of the People's Republic of China. (2025). *Outline for building an education power (2024—2035)*.