Review

The Impact Mechanism of Algorithmic Bias in Educational Recommendation Systems on Educational Process Fairness from an Algorithm Governance Perspective and the Construction of a Four-Dimensional Audit Framework

Xueqi Li[1,*]

[1] School of Education, Xizang University, Lhasa, China
* Correspondence: Xueqi Li, School of Education, Xizang University, Lhasa, China

Abstract: From an algorithm governance perspective, this study integrates the sociology of technology and educational equity theory. Utilizing literature review, logical deduction, and framework construction, it analyzes the generative logic and manifestations of algorithmic bias in educational recommendation systems, and reveals how such bias affects educational opportunities. The findings indicate that bias originates from skewed data collection, flawed model design, reinforcing interaction feedback, and weak institutional constraints. Through a triple mechanism of "copying, reinforcing, and solidifying", this bias creates "digital red line" barriers in resource accessibility, learning path planning, and developmental opportunity acquisition, thereby exacerbating intergroup educational inequality. Accordingly, a four-dimensional algorithm auditing framework is constructed, covering technology, process, outcome, and guarantee, with four specific auditing modules: data quality, model fairness, educational impact, and governance mechanisms. A multi-party collaborative governance pathway is also proposed. This research provides theoretical support and practical tools for mitigating risks to educational process fairness in the intelligent era, and promotes the transformation of educational recommendation systems from instruments of "personalized empowerment" to enablers of "fairness empowerment".

Keywords: algorithm bias; fairness in educational process; educational recommendation system; algorithm governance

Received: 09 July 2026
Revised: 13 August 2026
Accepted: 23 August 2026
Published: 27 August 2026

1. Introduction

Education development has been prioritised, systematically advancing the construction of an education powerhouse and promoting historic achievements and structural transformations in education in the new era [1]. In 2022, the Ministry of Education of the People' s Republic of China launched an action plan for educational digitalization aimed at comprehensively advancing the digital transformation of the education sector. This plan established a national smart education platform and created a mechanism for sharing high-quality educational resources across different educational fields [2]. As of September 2025, the national smart education platform has become the largest digital centre for educational resources in the world, with 170 million registered learners, and its functional modules have been continuously improved. The integration of artificial intelligence and education has reached an unprecedented depth and breadth [3]. Against this background, the forms of teaching and learning are undergoing profound changes. Continuing to promote the integration of artificial intelligence and education is not only a key measure for implementing the educational digitalization strategy, but also contributes to the cultivation of outstanding innovative talent and the enhancement of national technological competitiveness, thereby providing strong support for the construction of a high-quality education system and the achievement of educational modernization [4].

The educational recommendation system, as a core application of the integration of artificial intelligence and education, provides personalised learning path recommendations, intelligent resource allocation, and content push services. It recommends appropriate course content, practice exercises, or extracurricular activities to students, helping them autonomously select suitable growth paths and enhancing their learning experience. Such systems are expected to help bridge the gap in access to high-quality educational resources and to promote fairness in the educational process.

However, hidden risks lie behind this technological empowerment. Although algorithms function as technical tools, they are created by humans and are therefore not value neutral. Their decision-making processes embed social structural biases and technical design flaws, which give rise to algorithmic biases [5]. At present, although the application of artificial intelligence in education has injected new impetus into promoting educational equity, algorithmic bias in educational recommendation systems poses a potential threat to fairness in the educational process. These biases stem mainly from limitations in training data, imbalances in sample distribution, and systematic biases in data sources. As a result, artificial intelligence systems may unconsciously absorb and reinforce such biases, generating inequalities in key areas such as educational evaluation and resource allocation. In the context of personalised learning resource recommendation, the existence of algorithmic bias may result in some student groups receiving learning resources of lower quality or more limited coverage, thereby exacerbating imbalances in the distribution of educational opportunities and restricting the overall improvement of educational quality [6].

In 2023, UNESCO released the Guidelines for Generative Artificial Intelligence in Education and Research, calling on governments to regulate the application of generative artificial intelligence in education through the formulation of regulations and teacher training, and to establish a policy framework consistent with ethical requirements [7]. The guidelines warn that if the application of artificial intelligence in education lacks effective regulation, it may replicate or even amplify existing social inequalities, posing a serious challenge to educational equity. In practice, the bias risks of educational recommendation systems have already begun to emerge. Rural students may be directed towards low-quality resources due to insufficient data collection, students from disadvantaged families may be marginalised in resource allocation, and learners with special needs may struggle to obtain appropriate support due to design neglect. As a result, fairness concerns in education have gradually shifted from imbalances in physical resource distribution to unfairness in the educational process within digital spaces [8].

Educational equity is a fundamental component of social fairness and justice, and it is central to individual development opportunities and the sustainable development of society. China Education Modernization 2035 clearly states the need to focus on improving educational quality and promoting educational equity [9]. This study focuses on educational process equity in the digital domain, with particular attention to resource acquisition, learning paths, and the distribution of development opportunities, rather than treating equality of starting points as the central analytical concern. The educational recommendation system, with its powerful information filtering and agenda-setting capabilities, functions as a digital gatekeeper, directly influencing students' learning choices and development trajectories and thereby affecting educational process equity. If educational recommendation systems exhibit algorithmic bias or data discrimination, they may amplify existing educational imbalances. Without appropriate intervention, this may exacerbate disparities in access to high-quality educational resources among different groups and further widen the digital divide between regions and schools.

Global attention to algorithm governance has also been increasing steadily. In August 2024, the European Union' s Artificial Intelligence Act officially came into effect. This Act clarifies the framework for AI governance, emphasising that the assessment of AI learning outcomes constitutes a high-risk application and highlighting the importance of algorithmic fairness, transparency, and accountability systems [10]. In 2017, the State Council of the People' s Republic of China issued the New Generation Artificial Intelligence Development Plan, which stated that the wide application of artificial intelligence in education and other fields will greatly improve the precision of public services and comprehensively enhance people's quality of life. However, the Plan also cautioned that artificial intelligence is a disruptive technology with wide-ranging implications, which may bring about changes in employment structures, affect laws and social ethics, infringe upon personal privacy, and have profound impacts on government management, economic security, social stability, and even global governance. While vigorously developing artificial intelligence, it is therefore necessary to attach great importance to the potential security risks and challenges it may bring, strengthen forward-looking prevention and constraint guidance, and minimise risks in order to ensure the safe, reliable, and controllable development of artificial intelligence [11].

From an algorithm governance perspective, the governance of algorithmic bias cannot rely solely on technical repairs. It is necessary to connect the entire chain of technical detection, process supervision, outcome evaluation, and institutional guarantee. Against this background, relying solely on algorithmic optimisation is insufficient to eliminate educational inequality in the digital space. This study addresses three core questions from an algorithm governance perspective. First, through which stages does algorithmic bias in educational recommendation systems emerge, and through what internal mechanisms does it replicate, reinforce, and solidify inequality at the level of the educational process? Second, what dimensions, indicators, and closed-loop implementation processes should an algorithm auditing framework include if it is to integrate both technical standards and educational values and anchor the goal of educational process equity? Third, how can a collaborative governance system involving technology, institutions, and multiple stakeholders be constructed on the basis of auditing tools to prevent the erosion of educational process equity by algorithmic bias?

This study is grounded in the realities of educational digital transformation. It analyzes the generative logic and pathways of algorithmic bias in educational recommendation systems, extends the analytical dimension of educational process fairness into the digital domain, reveals the internal mechanism of copying, reinforcing, and solidifying through which algorithmic bias reshapes educational opportunities, and enriches the theoretical understanding of educational fairness in the intelligent era. Existing research on algorithmic fairness in education has mostly focused on bias detection at the technical level, with fewer studies constructing specialised auditing systems that integrate educational values [6]. This paper introduces algorithm governance theory into the context of educational recommendation systems, thereby achieving an interdisciplinary integration of algorithm governance and educational assessment theory. It addresses the limitations of traditional research that separates technical risk analysis from institutional governance, provides an analytical framework for studying fairness in educational digital transformation, and promotes the local application and theoretical innovation of algorithm governance in the field of education.

By examining the specific manifestations and actual harms of algorithmic bias in educational recommendation systems, this paper offers clear points of reference for risk perception among educational administrative departments, technology development enterprises, schools, and teachers. The four-dimensional algorithm auditing framework constructed in this study includes clearly defined dimensions, indicators, and closed-loop implementation processes, which can guide relevant entities in carrying out algorithmic bias detection, assessment, and rectification, and improve the fairness and inclusiveness of educational recommendation systems. The proposed governance policies can provide a basis for decision-making by policymakers seeking to improve the algorithm supervision system in education and to regulate intelligent education applications, thereby contributing to the core goal of enabling technology to empower fairness in the educational process.

2. Literature Review

2.1. The Technological Orientation of Research on Algorithmic Fairness in Education

The deep penetration of artificial intelligence in the field of education has promoted the development of personalized learning and precise teaching, but the fairness crisis caused by algorithmic bias has gradually become the focus of academic attention. The existing research on algorithmic fairness in education presents a significant technological orientation feature. The studies mostly focus on the detection of bias at the technical level and algorithm optimization, but generally ignore the particularity of educational scenarios and the uniqueness of educational values. Wang et al. found through a systematic review of 57 research papers on algorithmic bias in educational AI from 2013 to 2023 that current research is mainly divided into three categories: conceptual theory research, application research in educational scenarios, and research involving algorithm detection. Among them, 68% of the studies focus on pure technical dimension bias identification (such as slice analysis, quantification of fairness indicators) and bias reduction algorithms (such as adversarial training, causal intervention), and only less than 15% of the studies involve the design of governance frameworks that integrate educational values [6].

Gardner et al. further pointed out that the existing research has a significant tendency of technological transplantationism. A large number of studies simply adopt general bias detection schemes from the computer field, which often deviate from the educational ontology and are difficult to adapt to the core value demands of "cultivating moral character" and "developmental evaluation" in educational scenarios [12]. Yu and Wang emphasized from the perspective of educational ethics that such technical solutions only focus on the balance of group differences at the result level, but ignore the educational core elements such as students' learning process, effort level, and development potential, and may fall into the "cognitive outsourcing" trap, violating the essence of educational fairness, which is to promote the all-round development and individual growth of each student [13]. Huang et al. also emphasized that such technical solutions only focus on the balance of group differences at the result level, but ignore the educational core elements such as students' learning process, effort level, and development potential, and may violate the essence of educational fairness [14].

2.2. Research Gaps and Value Shift in Algorithmic Audit of Education

Algorithmic audit, as a core governance tool to ensure algorithmic fairness, is still in the initial stage of research and application in the field of education. The existing research on algorithmic audit in education has two key gaps: Firstly, the technological orientation of the audit framework. Most studies equate audit with technical detection, limiting the audit to pure technical aspects such as data verification and model bias identification, lacking systematic auditing of the entire life cycle of the algorithm (design, deployment, application, iteration) [15]; secondly, the absence of educational value. Existing audit tools generally do not embed core elements such as educational fairness, educational values, and educational ethics into the audit process, making it difficult for the audit results to reflect the real fairness requirements of educational scenarios [16]. General algorithmic audit tools are more focused on detecting technical risks and less consider the educational attributes of educational activities, relying solely on technical indicators for auditing, which can only achieve formal-level algorithm compliance but is unable to measure whether the algorithm truly promotes educational fairness, easily leading to the "misalignment" phenomenon of "technically compliant but educationally ineffective".

The international academic community also pays attention to this research gap. Chinta et al. proposed the FairAIED framework, indicating that the fairness issue of educational AI not only concerns result equality but also involves value choices at the epistemological level, such as "whose knowledge is valued," and the current technology-oriented audit schemes precisely ignore this dimension [17]. In addition, algorithmic audit of education must break away from the pure technical perspective and deeply integrate educational evaluation theory with technical audit tools to achieve substantive fairness of educational algorithms, rather than formal technical compliance [10]. The EU' s "Artificial Intelligence Act" classifies AI learning outcomes assessment as a high-risk application, emphasizing algorithm fairness, transparency, and accountability systems, further highlighting the urgency of establishing a specialized auditing system that integrates educational values [18]. Yu and Wang demonstrated that even if statistical-level fairness is achieved through technological means, if the auditing framework does not incorporate educational value judgments, it could still lead to the algorithmic solidification of educational inequality and even the emergence of new educational exclusion mechanisms [13].

2.3. Construction Path of an Algorithmic Audit System Integrating Educational Values

To break through the limitations of existing research and build a specialized auditing system that integrates educational values, three-dimensional transformations are required: from technical detection to the expansion of full life-cycle auditing, from general standards to adaptation to educational scenarios, and from a single entity to a multi-party collaborative governance model.

In the context of full life-cycle auditing, a closed-loop auditing mechanism of "data collection-algorithm design-application deployment-result feedback" needs to be established, embedding educational value judgments in each stage. For example, in the data collection stage, audit the educational representativeness of the training data to avoid algorithmic discrimination due to data bias. In the algorithm design stage, incorporate educational equity indicators to ensure that algorithmic decisions conform to educational laws. In the education scenario adaptation aspect, specialized auditing indicators should be constructed based on educational evaluation theories, such as converting "learning gain", "development potential", and "education opportunity equality" into quantifiable and auditable specific indicators. In the multi-party collaborative governance aspect, an auditing mechanism involving education workers, technical experts, ethicists, students, and parents should be established to ensure that the auditing process is both in line with technical norms and reflects educational value consensus.

Wang et al. call for future research to break away from a purely technical perspective, incorporate educational equity value into governance tool design, and develop specialized auditing tools adapted to educational scenarios, achieving the organic unity of technical rationality and educational value [6]. Jiang et al. have constructed an algorithmic auditing system integrating educational values, requiring the embedding of educational values in the auditing logic, implementing an auditing process covering the entire life cycle of the algorithm, promoting the participation of multiple entities such as managers, teachers, and technical personnel, and establishing a dynamic optimization and iteration mechanism [18]. Through the collaborative efforts of various mechanisms, it ensures that educational algorithms balance operational efficiency and fairness, achieving technological innovation and risk controllability [19]. Wang' s international comparative study emphasizes that building an educational algorithm auditing system requires drawing on international governance experience, while basing it on China' s own educational values, establishing fair auditing standards with cultural adaptability [5].

3. Core Concept Definition and Theoretical Foundation

3.1. Core Concept Definition

The four core concepts of this study have a clear internal logical relationship: algorithm bias is the source of risk, the educational recommendation system is the technical carrier of the risk occurrence, educational process fairness is the value evaluation criterion of this study, and algorithm auditing is the governance tool to resolve bias risks. Together, these four elements constitute a complete analytical framework for analyzing the fairness risks of educational recommendation systems.

3.1.1. Algorithm Bias

In recent years, algorithmic bias has become an interdisciplinary issue in fields such as education, law, information science, and computer science [20]. The academic community has not yet reached a completely unified definition. Some studies equate algorithmic bias with algorithmic discrimination, while others distinguish between the two: bias is the inherent bias at the system level, and discrimination is the empirical harm caused to specific groups after the bias is externalized [5].

In the context of educational recommendation system research, algorithm bias refers to the systematic unfairness tendency of educational artificial intelligence systems in the education content recommendation, learner assessment, and resource allocation processes, which are influenced by multiple factors such as training data, model design, interaction feedback, and institutional environment. Algorithm bias is a pre-condition risk condition. When this bias actually damages the development opportunities of learners, it transforms into algorithmic discrimination. Preventing the evolution of algorithm bias into algorithm discrimination is also an important starting point for conducting educational algorithm auditing [21].

3.1.2. Educational Recommendation System

The recommendation system originated in the field of information filtering, with its core function being to alleviate the information overload problem in the Internet environment [22]. With the penetration of big data and artificial intelligence into the education field, recommendation technology has gradually migrated to the educational scenario. The educational recommendation system relies on educational big data and machine learning algorithms to provide personalized learning resources, learning paths, and push services for students and teachers. Its ideal goal is to achieve individualized teaching, but it may also replicate the social structural inequalities embedded in the training data [23]. The research object of this study is limited to educational recommendation systems related to learners' learning resources and learning paths, excluding recommendation systems for teaching assistance and campus administrative management.

3.1.3. Educational Process Fairness

Educational fairness is generally divided into three levels: starting point fairness, process fairness, and outcome fairness [8]. Educational process fairness focuses on the resource supply, treatment, and development opportunities received by learners throughout the process of receiving education, emphasizing that regardless of the differences in family, region, and identity background of learners, they should be treated equally in teaching interaction, resource acquisition, and development opportunity allocation.

With the popularization of smart education, algorithms have become an important participant in the digital education process. The connotation of educational process fairness is no longer limited to the teacher-student interaction in offline classrooms, but also extends to equal opportunities in digital scenarios such as resource push and path planning dominated by algorithms [24]. This study focuses on educational process fairness in the digital domain and does not conduct a key analysis of starting point fairness.

3.1.4. Algorithm Auditing

Algorithm auditing was first introduced by Sandvig et al. in public algorithm research, initially used to detect the racial and gender group biases hidden in recruitment and credit automated decision-making systems [15]. Algorithm auditing is an assessment activity that comprehensively uses various means, such as technical detection, document review, user research, and compliance verification to evaluate the data quality, model logic, output consequences, and ethical compliance of automated decision systems, and is divided into internal auditing and third-party external auditing.

After algorithm auditing is migrated to the educational scenario, it cannot merely focus on pure technical indicators detection. It also needs to introduce educational value standards and take educational process fairness as an important evaluation criterion to achieve dual reviews of technical compliance and educational value [23].

3.2. Theoretical Foundation

3.2.1. Education Equity Theory

Zhou proposed an "four equalities and one tilt" analytical framework for educational equity, advocating that educational equity should ensure that every member of society enjoys equal educational rights, equal educational opportunities, equal public educational resource services, and equal opportunities for academic achievement development [23]. At the same time, it is necessary to implement compensatory tilt towards disadvantaged groups and decompose educational equity into three levels: starting point equity, process equity, and outcome equity. The theory of educational equity sets the core value scale for this research on auditing work. The algorithm auditing of the recommendation system cannot only focus on technical performance indicators, but also needs to prioritize the assessment of whether the system maintains the resource acquisition, path selection, and development opportunities of different groups in the digital education scenario. The audit indicator system needs to set compensatory fairness observation points for disadvantaged groups and anchor the core goal of educational process fairness [25].

3.2.2. Algorithm Governance Theory

Algorithm governance is an interdisciplinary research field spanning law, management, and sociology. This theory holds that algorithm bias is not solely caused by code defects, but is a product of the coupling of multiple factors such as technology, society, and institutions; bias governance cannot rely solely on technical repairs at the model level; it should cover the entire life cycle of the algorithm and coordinate the efforts of technical developers, regulatory agencies, and user groups to carry out governance [26-28]. The algorithm governance theory supports the overall thinking of the full life cycle auditing in this research and explains why auditing cannot be limited to static detection of the model, and it is necessary to set a system guarantee module, attract multiple stakeholders to participate in the auditing process, and not leave algorithm bias governance entirely to the independent completion of technology enterprises.

3.2.3. Education Evaluation Theory

Messick' s theory of consequence validity emphasizes that the effectiveness of assessment tools cannot be judged solely based on measurement accuracy, but must also fully consider the social impact and educational consequences brought about by the assessment practice; Starfield' s CIPP assessment model includes four modules: background, input, process, and outcome, advocating a balanced approach to both process evaluation and outcome impact evaluation. The above education evaluation theories jointly provide methodological support for the "technology-process-outcome-ensuring" four-dimensional auditing framework proposed in this research. The technical dimension corresponds to the input conditions of the system, the process dimension corresponds to the dynamic operation process of the system, the outcome dimension corresponds to the real educational consequences brought about by the algorithm, and the guarantee dimension corresponds to the external institutional background, so as to achieve an integrated assessment of the technical performance and educational value consequences of the recommendation system.

4. The Generation Path and Manifestation Forms of Algorithm Bias in Education Recommendation Systems

4.1. The Generation Path of Algorithm Bias

The generation of algorithmic bias in education recommendation systems is not the result of a single factor, but rather the result of multiple stages and factors working together throughout the entire lifecycle of the algorithm. It can be summarized into four core paths: data collection, model design, interaction feedback, and institutional constraints.

4.1.1. Data Collection Path

Data is the "fuel" of the algorithm. The quality of data directly determines the fairness of the algorithm' s output. The bias in the data collection process mainly stems from three aspects: first, the insufficiency of sample representativeness. The training data for education recommendation systems mostly comes from regions, schools, or groups with high informationization levels. Rural, remote areas, and disadvantaged students, due to the lack of equipment, network, or usage opportunities, have their data underestimated or missing, making it difficult for the algorithm to reflect their true needs and characteristics; second, the embedded historical bias in the data. Educational data carries historical unequal information, such as the resource gap between urban and rural areas and the academic differences between social classes, and directly using it for training will allow the algorithm to replicate and amplify the bias, such as the admission recommendation based on historical grades, which systematically recommends lower-level schools due to the lower average score of rural students [29]; third, the single dimension of data collection. Some systems only focus on explicit data such as grades and learning behaviors, ignoring implicit data such as family background, leading to a one-sided judgment of students' needs and the breeding of bias [30].

4.1.2. Model Design Path

Model design is the core aspect where algorithmic biases arise. The bias mainly stems from the cognitive limitations, value orientation, and technical choices of the designer. First, the subjective bias embedded in the algorithm designer's personal experience, value orientation, and cognitive limitations will unconsciously be integrated into the model design, causing the system to ignore the needs of specific groups, thereby forming design bias [31]. Second, the deviation of algorithm goals. Some recommendation systems use user activity, click rate, etc. as core optimization targets instead of educational equity or students' development goals, leading the algorithm to prioritize the resources that can improve click rates, while ignoring the real learning needs of disadvantaged groups, exacerbating the learning gap [32]. Third, the limitations of technical selection. Different algorithm models have different fairness characteristics. If the selection is inappropriate, it may lead to bias [33].

4.1.3. Interaction Feedback Path

The closed loop of interaction and feedback between users and the recommendation system may reinforce the initial bias, resulting in the "bias solidification" effect. First, the "echo chamber" and "filter bubble" effect. The system pushes similar content, narrowing the learning perspective. Disadvantaged groups, due to the low initial quality of resources, fall into a "low-energy cycle"; second, the selectivity bias of user behavior. The dominant group effectively uses the system and optimizes the recommendation based on user behavior, while disadvantaged groups, due to insufficient digital literacy, have difficulty providing feedback, exacerbating the bias in demand identification; third, social identity reinforcement. The system' s labeling of groups affects self-perception and others' expectations, forming a "label-behavior-result" vicious cycle, restricting development opportunities [34].

4.1.4. Institutional Constraint Path

The absence or inadequacy of institutional constraints provides space for the generation of algorithmic bias. First, incomplete laws and regulations. Currently, there are not yet complete laws and regulations in China regarding algorithm bias in the education field. There is a lack of clear norms for algorithm design, data use, and result evaluation, and enterprises lack fairness constraints [35]. Second, there is an absence of industry standards. The fairness assessment standards and technical specifications for education recommendation systems have not yet been formed, making it difficult to ensure the fairness of the system. Third, the absence of regulatory mechanisms. The government' s supervision of education recommendation systems is insufficient, making it difficult to promptly identify and rectify algorithm bias issues [15]. Fourth, the lack of complaint, appeal, and relief channels. If students, parents, or teachers encounter unfair treatment due to algorithm bias, they lack effective complaint, appeal, and relief channels, making it difficult to protect their rights and interests.

4.2. Manifestation Forms of Algorithm Bias

In the context of educational recommendation systems, algorithmic biases mainly manifest in four forms: resource allocation bias, path planning bias, labeling bias, and opportunity deprivation bias. These forms interweave with each other, jointly influencing the allocation of educational opportunities.

4.2.1. Resource Allocation Bias

This refers to the differential treatment of different groups in the recommendation system when pushing learning resources, resulting in the disadvantaged groups obtaining resources of lower quality, quantity, and adaptability compared to the advantaged groups. Specifically, in terms of urban-rural differences, rural students mostly receive basic repetitive content and lack innovative and expanded resources, while urban students can obtain high-quality resources such as famous teacher courses and personalized tutoring; in terms of class differences, students from disadvantaged families have difficulty receiving paid high-quality resource recommendations, while students from advantaged families continuously receive related recommendations; in terms of group differences, special-needs students only receive general resources and have difficulty meeting their individualized needs.

4.2.2. Path Planning Bias

Algorithmic biases lead to the recommendations of the recommendation system in terms of learning paths, academic advancement directions, and career planning, restricting the development space of disadvantaged groups and reinforcing the development trends of advantaged groups. Specifically, the learning paths are fixed. The algorithm, based on initial grades or family backgrounds, plans "low-level, low-expected" paths for disadvantaged groups and "high-level, high-expected" paths for advantaged groups; academic advancement recommendations are biased, influenced by historical data or regional biases, recommending local lower-level schools to rural students and high-quality schools in other regions to urban students; career planning is limited, with labels based on gender, family background, etc., such as recommending traditional occupations like education and nursing for females and high-paying occupations like technology and finance for males.

4.2.3. Labeling Bias

This refers to the recommendation system applying fixed labels to students through data analysis and differentiating recommendations based on these labels, reinforcing stereotypes and restricting development. This includes: static labels being fixed, the algorithm applying labels based on data at a single point in time, ignoring the dynamic development and potential of students, such as "learning-challenged students" continuously receiving low-difficulty content, making it difficult to improve; negative labels being strengthened, the algorithm being more sensitive to the negative behaviors of disadvantaged groups, frequently pushing critical feedback and low-expected content, affecting self-identity and learning motivation; group labels discrimination, binding specific groups with negative characteristics, forming systematic group discrimination.

4.2.4. Opportunity Deprivation Bias

Algorithmic biases directly lead to the loss of important educational opportunities for disadvantaged groups, such as competition participation, scholarship application, and admission to high-quality schools. Specifically, competition recommendations are lacking, the system prioritizing development opportunities like competitions and research projects for advantaged groups based on historical grades or family backgrounds, excluding disadvantaged groups; admission recommendations are unfair, the college admission recommendation system being influenced by regional, family background, etc. biases, reducing the compatibility score for rural students and disadvantaged families, and reducing their admission opportunities.

5. Algorithm Bias Reshapes the Core Mechanism of Educational Opportunities

Algorithm bias reshapes educational opportunities through the "copy-reinforce-solidify" triple mechanism, creating a vicious cycle of "technology exacerbating social inequality".

5.1. Copying Mechanism

Algorithm bias replicates social structural inequalities. The training data of the education recommendation system embeds historical educational inequality information, such as the urban-rural gap, class gap, racial discrimination, etc. Without effective intervention, the algorithm will directly replicate these structural biases and convert them into differentiated recommendation results. This replication extends social structural inequalities from "offline" to "online".

5.2. Reinforcing Mechanism

Algorithm bias reinforces educational inequality through interactive feedback. The "filter bubble" and "echo chamber" effects of the recommendation system make students' learning environments continuously narrow. Vulnerable groups, due to the lower quality of the initial resources they obtain, continue to receive lower-quality resources in the subsequent push, and the learning gap continues to widen; while the advantaged groups continuously receive high-quality resources, and their learning advantages continue to strengthen. At the same time, the tagging behavior of the algorithm reinforces the stereotypes of society towards specific groups, affecting students' self-identity and learning motivation, further exacerbating the learning gap and accelerating the development process of inequality.

5.3. Fixing Mechanism

Algorithm bias solidifies educational inequality through institutions and culture. The long-term existence of algorithm bias will form a stable distribution model of educational opportunities, which is gradually accepted and internalized by society, forming a "digital red line" type of institutional discrimination. At the same time, algorithm bias will affect the formation of educational culture, strengthening the "technology determinism" cognition, making people pursue educational equity as a value pursuit, and consider algorithm recommendation results as an inevitable choice of "objective and fair", further solidifying educational inequality. This fixing mechanism makes educational inequality shift from "temporary" to "long-term", from "changeable" to "difficult to reverse", posing a fundamental threat to educational equity.

In conclusion, the education recommendation system, through the "copy-reinforce-solidify" triple internal mechanism, replicates, amplifies, and further solidifies the existing educational inequality in the real world into the digital learning scenario, forming a vicious cycle of "technology exacerbating social inequality". Simply relying on technical means at the algorithm model level to correct the bias is difficult to break this chain from the root. A specialized algorithm auditing tool that takes into account technical detection, operation process monitoring, educational outcome assessment, and institutional constraints must be established. The auditing framework needs to set corresponding observation points for each of the three mechanisms: identifying data and model biases brought by the copying mechanism at the source, intervening in the bias reinforcement effect brought by interactive feedback during the system operation stage, and cracking the risk of long-term solidification of inequality through the institutional guarantee module. Based on this, a "copy-reinforce-solidify" four-dimensional algorithm auditing framework suitable for the education recommendation system scenario is constructed in the following text.

6. Construction of the Algorithm Audit Framework for Education Recommendation Systems

Based on the analysis of the generation path of algorithm bias and the "copy-reinforce-solidify" mechanism of action in the previous section, combined with the theories of educational equity, algorithm governance, and educational assessment, this study constructs a four-dimensional algorithm audit framework of "technology-process-result-guarantee". This framework aims to maintain the fairness of the educational process in the digital field and takes into account both technical rationality and educational values, forming a fairness assessment system that runs through the entire life cycle of the algorithm.

6.1. Principles and Goals of the Audit Framework Construction

6.1.1. Construction Principles

The construction principles of the audit framework mainly consist of five aspects: first, the educational orientation principle. The audit framework should take ensuring educational equity and promoting students' all-round development as the core goal, while also considering technical performance and educational value [36]; second, the full life cycle principle. The audit needs to cover the entire life cycle of the algorithm, including design, data collection, model training, operation iteration, and result output, to achieve full-process supervision; third, the principle of combining scientificity and practicality. The audit indicators should be scientific and reasonable, and possess measurable and operable characteristics, facilitating practical application; fourth, the principle of multi-party participation. The audit process should involve multiple stakeholders, such as government, enterprises, schools, teachers, students, and parents, ensuring the objectivity and comprehensiveness of the audit process; fifth, the principle of dynamic adjustment. The audit framework needs to be continuously optimized and improved based on technological development, changes in educational policies, and practical feedback.

6.1.2. Construction Goals

First, identify algorithm bias. Precisely identify the latent bias risks in all stages of the algorithm recommendation system' s life cycle, clarify the manifestation, scope of influence, and degree of harm of the bias; second, assess the level of fairness. From multiple dimensions such as technical foundation, operation process, educational actual consequences, and institutional conditions, comprehensively judge the fairness level of the system, and conduct a substantive assessment of the fair implementation of the educational process; Third, output rectification basis. For the identified risk points in the audit, form targeted optimization and rectification suggestions; fourth, establish a long-term mechanism. Promote the formation of an "audit-rectification-evaluation-recheck-optimization" closed loop to achieve continuous governance of the fairness of the education recommendation system.

6.2. Four-Dimensional Audit Framework of "Technology-Process-Result-Guarantee"

This study' s four-dimensional audit framework corresponds to four major levels: algorithm source technology conditions, dynamic interaction operation, actual educational output consequences, and external institutional guarantees. The following table summarizes the audit objectives, key observation points, and the sources of bias risks that this dimension needs to respond to.

Based on a comprehensive review of relevant knowledge, such as educational assessment theory and algorithm governance theory, this paper constructs a four-dimensional algorithm audit framework of "technology-process-result-guarantee" , forming a fairness assessment system that runs through the entire life cycle of the algorithm. This framework includes four major modules, twelve first-level indicators, and thirty-six second-level indicators, which can comprehensively cover the key points of fairness throughout the entire life cycle of the algorithm, providing solid system support for the scientific prevention and control of algorithm bias, and offering a systematic and structured analysis path for a comprehensive examination of algorithm fairness.

First, the technical dimension audit. The technical dimension audit starts from the core of algorithm design, focusing on the fairness basis in the data, model, and implementation stages. Its core task is to identify and control potential bias sources during the algorithm construction stage, including the assessment of data representativeness, the rationality of model logic, and the inclusiveness of system implementation. The audit methods include data sampling, explainability technical analysis, etc., striving to build a fair defense line at the technical level.

Second, the process dimension audit. This dimension extends the perspective to the dynamic operation process of the algorithm, examining the potential fairness deviations that may occur in the application interaction. It focuses on the consistency of algorithm decision logic, the effectiveness of the user feedback mechanism, and the rationality of dynamic updates, aiming to prevent the algorithm from solidifying or amplifying existing biases during continuous operation.

This dimension emphasizes real-time monitoring, comparative analysis, and iterative evaluation to form a process-based examination of the algorithm' s behavior.

Next is the result dimension audit. Based on the actual impact of the algorithm' s output, it particularly focuses on the differentiated educational outcomes generated among different groups. This dimension focuses on three aspects: resource allocation, learning development, and social impact. Through quantitative analysis, it assesses the actual effects of the algorithm on educational opportunities, individual development paths, and the social fairness landscape. The audit reveals the substantive impacts brought about by algorithmic biases by tracking result data and conducting impact assessments.

Finally, the guarantee dimension audit is conducted from the governance and institutional perspective, aiming to build a long-term mechanism to support algorithmic fairness. Its focus is on reviewing the completeness and effectiveness of relevant laws and regulations, regulatory systems, and responsibility arrangements to ensure that the algorithm is always under institutionalized fair constraints throughout its entire life cycle. This dimension promotes the formation of a systematic guarantee with clear responsibilities, strong supervision, and effective remedies through institutional analysis, cobaltamine assessment, and case review.

6.3. Audit Implementation Process and Methods

6.3.1. Audit Implementation Process

The algorithm audit of the educational recommendation system follows a closed-loop process of "preparation-implementation-evaluation-rectification-review" .

1. Preparation stage: Clearly define the audit object, scope, and objectives; form a diverse audit team composed of technical experts, education experts, ethical researchers, and representatives of teachers and students.
2. Implementation stage: Use a variety of auditing methods to collect system technical materials, operational data, and user experience information, and conduct itemized audits from four dimensions: technology, process, result, and guarantee.
3. Evaluation stage: Integrate multi-dimensional auditing information, assess the risk level of algorithmic biases, and compare against the value standards of educational process fairness to form a formal audit report, listing risk points and causes of problems.
4. Rectification stage: Feedback the audit conclusions to the system development and operation entities, propose feasible rectification and optimization suggestions, and track the progress of rectification.
5. Review stage: Conduct a re-audit after rectification to verify the actual effect of the rectification and form closed-loop management.

6.3.2. Core Audit Methods

The commonly used auditing methods in the audit process include the following five. First, the technical detection method, using interpretable AI, data mining, system testing, and other technical tools to detect data sample deviations and model decision logic, and identify potential bias hazards at the technical level; second, the data analysis method, capturing the actual operation logs of the system, resource push records, and user behavior data, and statistically comparing the differentiated output results of resources, paths, and development opportunities among different regions and family background groups; third, the user research method, collecting subjective perceptions of fairness of recommendations from students and teachers through questionnaires, interviews, and focus groups, to compensate for the insufficiency of pure technical detection in capturing subjective educational experiences; fourth, the case study method, selecting typical learner cases to track the actual impact of algorithm recommendations on individual learning opportunities, and revealing the actual path of algorithmic biases' effect; fifth, the compliance and expert assessment method, conducting compliance verification by comparing with national policies, industry norms, and regulatory experiences from abroad, and inviting experts in education and algorithm ethics to review the audit results.

7. Strategies for Algorithm Bias Governance and Education Equity Assurance

Drawing on the four-dimensional auditing framework of technology, process, result, and guarantee, this section proposes an integrated governance system oriented towards the interlocking dimensions of technology, institution, culture, and capability. Technical optimisation is designed to address bias at its source: data collection mechanisms should be broadened to include rural, remote, and disadvantaged learner populations, with periodic bias detection, cleaning, and quality control; model design should incorporate fairness constraints, interpretability, and group-specific testing, rather than optimising solely for click-through rates; and interactive feedback should be diversified through exposure to resources beyond learners' existing interests, accessible feedback channels, user rights to adjust unreasonable recommendations, and dynamic strategy correction. Together, these measures aim to prevent and correct bias across the data, model, and interaction loop.

Institutional construction and multi-party governance are mutually reinforcing. At the regulatory level, laws, regulations, and industry standards should be refined for high-risk educational AI, drawing on instruments such as the EU Artificial Intelligence Act; a regular audit and supervision mechanism should embed the four-dimensional framework within government-led, third-party-implemented audits, combining pre-launch review with post-launch monitoring; and complaints, accountability, and remediation procedures should be clarified to safeguard learners' rights. Governance responsibilities should be distributed across government, enterprises, schools, teachers, students, and independent auditors, supported by collaborative platforms and self-regulatory guidelines. International cooperation should facilitate the localization of foreign standards, participation in global standard-setting, joint technical research, and governance of cross-border recommendation systems.

Digital literacy enhancement addresses the human capability to identify and respond to algorithmic bias. Student-centred measures include integrating algorithmic literacy into school curricula, providing targeted digital assistance for rural and disadvantaged learners, and fostering critical evaluation of algorithmic labels. Teacher training should develop the capacity to identify, intervene in, and report biased recommendations, thereby generating audit-relevant feedback. Public understanding should be strengthened through media campaigns, publication of typical bias cases, reporting platforms, and greater algorithmic transparency. These measures complete the technology--institution--culture--capability system and collectively support educational process fairness.

8. Research Conclusions and Prospects

This study has examined the mechanisms through which algorithmic bias in educational recommendation systems reshapes digital educational opportunities. Four conclusions emerge. First, algorithmic bias arises from the interaction of data collection, model design, interactive feedback, and institutional constraints, and manifests as resource allocation bias, path-planning bias, labelling bias, and opportunity deprivation. Second, algorithmic bias operates through a "copy--reinforce--solidify" mechanism: it replicates existing offline educational inequalities, amplifies group learning gaps through interaction feedback, and consolidates inequality under institutional and cultural conditions. Third, the technology--process--result--guarantee auditing framework provides a structured tool for identifying bias across technical sources, dynamic operation, educational outcomes, and external safeguards. Fourth, effective governance requires an integrated technology--institution--culture--capability system.

Nevertheless, the study has limitations. It is primarily a speculative framework-building exercise based on literature review and logical deduction; the proposed auditing indicators and their weights have not been validated through pilot testing or Delphi consultation. The analysis addresses educational recommendation systems in general and does not differentiate by educational stage, functional type, or scenario-specific auditing focus. Moreover, the absence of in-depth case studies of domestic platforms limits the micro-empirical evidence concerning learners' psychological motivations and self-identity.

Future research should therefore pursue three directions. The auditing framework should be empirically tested and refined through pilot audits at different educational stages, with expert consultation to optimise indicators. Micro-level mechanisms should be investigated further to clarify how algorithmic bias affects learning motivation and self-identity, thereby strengthening the empirical basis of the "copy--reinforce--solidify" thesis. Finally, the framework should be translated into practice by developing third-party audit procedures and detailed implementation rules that can inform educational administrative regulation.

References

1. T. T. Zhuang, X. J. Shang, and T. F. Qin, "Frontier exploration and future landscape of higher education in the intelligent era: A review of the parallel session on 'AI leading high-quality development of higher education' at the 2025 Global Smart Education Conference," *China Educational Informatization*, no. 11, pp. 37–47, 2025.
2. Ministry of Education of the People's Republic of China, "National smart education platform officially launched," [Online]. Available: https://www.moe.gov.cn/jyb_xwfb/s5147/202203/t20220329_611601.html. [Accessed: Mar. 29, 2022].
3. Ministry of Education of the People's Republic of China, "Press conference on the progress and achievements of accelerating the construction of an educational powerhouse during the '14th Five-Year Plan' period," [Online]. Available: https://www.cernet.edu.cn/xxh/focus/xs_hui_yi/202509/t20250924_2691631.shtml. [Accessed: Sep. 23, 2025].
4. Z. T. Zhu and J. Hu, "Essential analysis and research prospects of educational digital transformation," *China Educational Technology*, no. 4, pp. 1–8+25, 2022.
5. S. W. Jia and H. Yan, "A systematic review of the concept, philosophical foundations, and consequences of algorithmic bias," *Journal of Library Science in China*, vol. 48, no. 6, pp. 57–76, 2022.
6. Y. M. Wang, D. Wang, H. J. Wang, and C. C. Liu, "Algorithmic fairness: Logic and governance of algorithmic bias in educational artificial intelligence," *Open Education Research*, vol. 29, no. 5, pp. 37–46, 2023.
7. United Nations Educational, Scientific and Cultural Organization, "Guidance for generative AI in education and research," [Online]. Available: https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research. [Accessed: 2023].
8. S. T. Ma, G. J. Shi, and Q. Wang, "The phenomenon of artificial intelligence technology encroaching on educational equity and its remedies," *Distance Education in China*, vol. 45, no. 7, pp. 98–114, 2025.
9. Central Committee of the Communist Party of China and State Council, "China education modernization 2035," [Online]. Available: https://www.gov.cn/zhengce/2019-02/23/content_5367987.htm. [Accessed: Feb. 23, 2019].
10. European Parliament and Council of the European Union, "Regulation (EU) 2024/1689 on artificial intelligence (AI Act)," [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689. [Accessed: 2024].
11. State Council of the People's Republic of China, "Notice of the State Council on issuing the next generation artificial intelligence development plan," [Online]. Available: https://www.gov.cn/zhengce/zhengceku/2017-07/20/content_5211996.htm. [Accessed: Jul. 20, 2017].
12. J. R. Gardner, C. Brooks, and R. S. Baker, "Evaluating the fairness of predictive student models through slicing analysis," in *Proc. 9th Int. Conf. Learning Analytics & Knowledge*, New York, NY: ACM, 2019, pp. 225–234.
13. S. Q. Yu and F. C. Wang, "The cognitive outsourcing trap of artificial intelligence educational applications and its transcendence," *E-education Research*, vol. 44, no. 12, pp. 5–13, 2023.
14. R. H. Huang, G. L. Zhang, and M. Y. Liu, "Technological ethics orientation and risk regulation for smart education," *Modern Educational Technology*, vol. 34, no. 2, pp. 13–22, 2024.

15. L. Dai and Z. T. Zhu, "Governance of ethics and moral risks in educational artificial intelligence: Problem clarification and targeted measures," *Journal of the Chinese Society of Education*, no. 12, pp. 31–37, 2024.
16. S. V. Chinta et al., "FairAIED: Navigating fairness, bias, and ethics in educational AI applications," *arXiv preprint arXiv:2407.18745*, 2024.
17. S. Messick, "Validity of psychological assessment: Validation of inferences from persons' responses' and performances as scientific inquiry into score meaning," *American Psychologist*, vol. 50, no. 9, pp. 741–749, 1995.
18. B. Jiang, Y. W. Ding, and Y. A. Wei, "The core technological engine of educational digital transformation: Trustworthy educational artificial intelligence," *Journal of East China Normal University (Educational Sciences)*, vol. 41, no. 3, pp. 52–61, 2023.
19. X. N. Wang, "Risks, challenges, and governance pathways of AI-empowered primary and secondary education: A five-country comparative study based on the Estonian case," *Journal of Comparative Education*, no. 3, pp. 125–138, 2026.
20. P. Yue and Y. Miao, "Social governance: Problems and regulation of algorithmic bias in the age of artificial intelligence," *Journal of Shanghai University (Social Sciences Edition)*, vol. 38, no. 6, pp. 1–11, 2021.
21. Z. Z. Ding, "Application and optimization of artificial intelligence algorithms in intelligent recommendation systems," *Software*, vol. 46, no. 6, pp. 40–42, 2025.
22. G. Chen, "Design of personalized educational resource recommendation system under big data," M.S. thesis, Shenzhen University, Shenzhen, China, 2019.
23. H. Y. Zhou, *On Educational Equity*. Beijing, China: People's Education Press, 2010.
24. I. D. Raji and M. Mitchell, "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in *Proc. 2019 AAAI/ACM Conf. AI, Ethics, and Society*, New York, NY: ACM, 2019, pp. 33–40.
25. X. D. Ding, "On the legal regulation of algorithms," *Social Sciences in China*, no. 12, pp. 138–159+203, 2020.
26. J. Guo and Q. Rong, "AI-empowered educational equity: International consensus, practical obstacles, and implementation pathways," *Journal of Southwest University (Social Sciences Edition)*, vol. 51, no. 2, pp. 247–258+315, 2025.
27. H. J. Xiao, "Algorithmic responsibility: Theoretical justification, panoramic portrait, and governance paradigm," *Management World*, vol. 38, no. 4, pp. 200–226, 2022.
28. H. Suresh and J. Guttag, "A framework for understanding unintended consequences of machine learning," *arXiv preprint arXiv:1901.10002*, Jan. 29, 2019.
29. J. Y. Bai, "Ethical governance of educational artificial intelligence: Practical challenges and implementation pathways," *Chongqing Higher Education Research*, vol. 12, no. 2, pp. 37–47, 2024.
30. Z. Liu, "Algorithmic discrimination: Manifestations, causes, and governance strategies," *People's Tribune*, no. 2, pp. 64–68, 2022.
31. S. Fazelpour and D. Danks, "Algorithmic bias: Senses, sources, solutions," *Philosophy Compass*, vol. 16, no. 8, p. e12760, 2021.
32. J. J. Qu, Y. C. Wang, and Y. D. Zhao, "Social experiments on artificial intelligence: Ethical norms and operational mechanisms," *China Soft Science*, no. 11, pp. 74–82, 2022.
33. A. Olteanu, C. Castillo, F. Diaz, and E. Kiciman, "Social data: Biases, methodological pitfalls, and ethical boundaries," *SSRN Electron. J.*, 2017.
34. P. Chen and Q. L. Yu, "Empowering the right to education through educational digitalization: Contemporary connotations, legal risks, and regulation," *Journal of the Chinese Society of Education*, no. 4, pp. 44–50, 2024.
35. D. Wu and Q. Guo, "Ethical challenges of intelligent technology-empowered teaching: Characterization, attribution, and mitigation," *Open Education Research*, vol. 30, no. 4, pp. 20–27, 2024.
36. Z. L. Cao, F. M. Yang, M. Hong, and S. J. Wu, "Construction and implementation path of university big data audit framework from the perspective of full-coverage auditing," *Internal Auditing in China*, no. 9, pp. 51–56, 2025.