



Research on the Implicit Intervention Mechanisms and Collaborative Governance of AI Algorithmic Bias in the Integration of Ideological and Political Education Across Universities, Secondary, and Primary Schools

Yan Shen and Yuanpeng Duan*

Chongqing Technology and Business University
Institute of Marxism, Chongqing, China
e-mail: 18213836606@163.com; 40643580@qq.com

Abstract. Artificial intelligence algorithm bias impacts the integrated system of ideological and political education in universities, primary and secondary schools through the triple mechanism of data source distortion, modeling fuzziness and system closed-loop discrimination. Erosion target consistency, synergy and effectiveness. Based on the Marxist view of science and technology and the theory of educational ecology, It reveals the internal logic of the framework of “ five-dimensional whole-person education ”. Governance path needs to build a system-technology-subject collaborative framework: to establish a dynamic data calibration mechanism to eliminate historical bias, Implement algorithm life-cycle supervision to constrain technical power, develop special detection tools to enhance transparency, And through the multi-agent cooperation network balance tool rationality and value rationality. The framework aims to solve the reconstruction risk of the algorithm to the educational value coordinate, and maintain the ideological security and educational fairness. To ensure that the integration system of ideological and political education in the digital age adheres to the fundamental task of “ establishing morality and cultivating people ”, and provides a technical governance paradigm for the modernization of education.

Keywords: Artificial intelligence, Integration of ideological and political courses, Algorithmic bias ,Hidden intervention, Collaborative governance

“AI and data science are indeed revolutionizing education, which has its positives and negatives as it becomes more integrated with the academic environment”¹, with large AI models and algorithms deeply embedded in various social domains, becoming fundamental infrastructure for societal operations and a core driving force for economic development. Behind this seemingly objective data-driven system lies a critical issue demanding attention - algorithmic bias. How to mitigate algorithmic bias, balance personalized recommendations with fairness in algorithm design and implementation, thereby ensuring the impartiality and diversity of algorithmic recommendations throughout the integrated education system of ideological and political courses across universities, secondary, and primary schools, constitutes an urgent

challenge for applying AI algorithms to the holistic development of ideological and political education.

1 The Interaction Mechanisms and Theoretical Framework between AI Algorithm Bias and the Integrated Collaborative Governance of Ideological-Political Education across Universities, secondary, and Primary Schools

Algorithmic bias refers to the phenomenon in which computer systems, throughout the entire process of research and development, data processing, and information dissemination, systematically deviate from objective facts and result in unfair treatment toward specific groups or information content. This deviation stems from the subjective stances of algorithm designers or biases inherent in the training data. Such systematic bias manifests at the technical implementation level and extends to the effects of information dissemination.

1.1 Generative Logic of Algorithmic Biases

Human interference and cognitive limitations lead to bias. Algorithm designers whose subjective value orientation and cognitive blind spots impact them inscribe into algorithmic rules those existing biases or stances existent in society or their personal life. Designers, for their part, do not maintain an objective, neutral decision-making process, which risks institutionalizing preexisting biases. Based on the algorithm, nearly 40 percent of students were downgraded as compared to their teacher's assessment. Students who attended public schools were disproportionately downgraded compared to their private school-attending peers, creating discrepancies based on class². Secondly, structural bias and distortion in data sources propagate bias. Raw data inherently carries societal structural biases, while the data collection process suffers from sampling imbalances, commercial fabrication, and the filtering of minority group data. Technical oversimplification during data cleaning, such as directly removing data with indistinct feature further exacerbates the vicious cycle of "biased input-biased output."³Thirdly, technical flaws in algorithmic operations perpetuate bias. Deep learning models face technical constraints during both training and deployment phases. Computational rules are easily contaminated by initial biased data, while machine learning overemphasizes statistical significance while neglecting long-tail phenomena, coupled with emergent deviations during algorithmic interactions. Computational constraints and the absence of feedback mechanisms hinder dynamic bias correction, resulting in a self-reinforcing "bias black box."⁴The flow chart of the mechanism of algorithmic bias formation is shown in Fig.1.

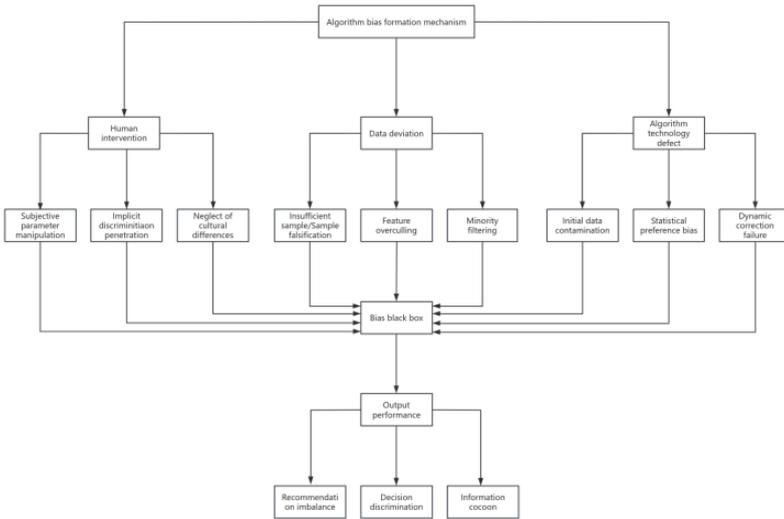


Figure 1. Flow Chart of the Algorithmic Bias Formation Mechanism

1.2 Core Features of the Integrated Ideological and Political Education Across Universities, Secondary, and Primary Schools

The political orientation and goal alignment are one of its key features. Both the ideological shaping emphasized by its political nature and the fundamental task of “fostering virtue through education” in practice reflect the essential attribute of ideological and political education courses being guided by political direction. Its core objective is to ensure consistent educational direction across different academic stages through vertical coherence and horizontal integration, ultimately serving the national strategy of cultivating socialist builders and successors, thereby forming a continuous value orientation throughout the educational process.

Because of its systemic coordination and structural hierarchy, horizontal consistency and vertical progression constitute the pillars for establishing a scientific education system. Horizontal consistency, on the other hand, needs greater unity of teaching philosophy and stricter logical coherence between knowledge: it should realize the synchronization of educational goals between different stages. Vertical progression focuses on the hierarchical promotion and dynamic internal articulation of teaching content, guiding the learners away from repetitive knowledge output and attaining stepwise improvement at different cognitive levels. Another essential feature is, at the same time, practical applicability and educational effectiveness. The practical application runs throughout the entire process, highlighting the move from education objectives to a real-life education outcome. These include both theoretical requirements for Marxist theory to be incorporated into actual educational practices, as well as practical goals for achieving knowledge transformation and innovative ability cultivation through curriculum integration. It reflects the linking preference of abstract political concepts with students’ perceptible and actionable value identification with

an accurate education at all levels of closed-loop shapes of into textbooks, into the classroom, into the mind the success of the outcome.

1.3 Theoretical Basis for the Collaborative Governance of AI Algorithm Bias and the Integration of Ideological and Political Courses in Universities, Secondary, and Primary Schools

Regarding Marxist views on science and technology. The Marxist perspective on science and technology reveals the dialectical logic of the relationship between technology and society. In terms of basic orientation, collaborative governance should be based on Marxist views on science and technology to provide a macroscopic value orientation. The critical theory of technological alienation reveals that algorithms driven by capitalist logic actually contain certain biases, which stem from the usurpation of educational value rationality by instrumental rationality. Ideological reproduction theory reveals that ideological and political education courses are the main channel for disseminating national ideology, and the integrated design of ideological and political education courses at different levels of education should be regarded as an educational chain that conveys values.

Regarding the theory of holistic human development. Collaborative governance should make ideological and political education's technological applications serve the educational goal of "Five Education Initiatives." Specifically, algorithmic transparency and maneuverability should be regarded as a prerequisite for strengthening ideological and political education. Otherwise, the realization of knowledge, values, and competency objectives will be affected, which is detrimental to the holistic realization of ideological reproduction educational goals. Educational systems science and cognitive development stage theory are also important theoretical resources for collaborative governance. On the one hand, they provide methodological support for collaborative governance, requiring adherence to educational laws. Educational systems science reveals that the ideological and political education system at different levels of primary, secondary, and higher education is like an ecological system with its own laws and order. The four theories of algorithmic biases easily disrupt the ecological balance of the integrated ideological and political education system across primary, secondary, and higher education. On the other hand, cognitive development stage theory reveals that students' thinking abilities have an irreversible development from concrete operational to formal operational. Algorithmic content pushing that ignores the stage characteristics of students' thinking abilities is easy to cause structural fractures in the value cultivation of students with value development in different stages. However, algorithmic biases risk fragmenting the collaborative network between "ideological-political courses" and "curriculum-based ideological-political education", necessitating governance mechanisms to rebuild technology-empowered systemic synergy. This theoretical dimension delineates the scientific boundary for governance: Algorithm design must deeply integrate with the characteristics of educational stages and the laws of cognitive development.

Furthermore, technology governance and algorithmic justice theory also constitute one of its theoretical foundations. Technology governance theory offers practical op-

erational tools for collaborative governance. The theory of responsible innovation advocates embedding ethical reviews at the algorithm development stage, requiring the integration of socialist core values through “value-sensitive design” at the technological source. Algorithmic justice theory proposes dual governance standards: Procedural justice demands establishing transparent mechanisms for educational algorithms, disclosing the logic of resource allocation and decision-making bases. Outcome justice requires employing impact assessment tools to quantitatively analyze the exclusionary effects of algorithmic biases on different groups. This theoretical dimension provides a practical path: through institutional regulation, technical improvement, the main body of the coordinated force, Realize the dynamic balance between instrumental rationality and value rationality.

2 The Implicit Intervention Mechanism of Artificial Intelligence Algorithm Bias on the Integration of Ideological and Political Education Across Universities, Secondary, and Primary Schools

“Artificial Intelligence has the potential to improve education through personalized learning, educational accessibility and inclusion, intelligent tutoring systems, automated assessments, and learning analysis”³. The implicit mechanisms of artificial intelligence algorithmic bias are intervening in the ecological system of integrated ideological and political education. The historical inertia in data collection, the marginalization of groups during modeling processes, and the closed-loop system operation collectively form a triple-layered implicit intervention mechanism of algorithmic involvement in ideological education integration: from structural distortion in educational data to the dissolution of cognitive model differences, ultimately evolving into systemic bias proliferation. This technological discipline not only obscures authentic demands within the educational field of integrated ideological and political courses but also accomplishes intelligent reconstruction of algorithmic bias in the digitalize process. “Bias, which can occur at all stages of the machine learning life cycle, is a multilayered phenomenon encompassing historical bias, representation bias, measurement bias, aggregation bias, evaluation bias and deployment bias”⁴. No matter how algorithm is, it inevitably has three hidden dimension of hidden harm on educational practice: although algorithm is merely a technical governance mechanism, it also would bear the path dependence in acquisition of data, homogenization tendency in model construction, self-reinforcing bias in operation – these three hidden dimension jointly shape educators’ practices, and create new form of digital alienation in pedagogical space.

2.1 Triple Distortions at the Data Source: Historical Inertia and Group Silencing

The data foundation of AI also carries three layers of bias. The first one is historical bias, which means that the algorithm embeds the social inequality into the data, and renders the current social inequality as the algorithm criterion, thus making the marginalized group’s social circumstances become a part of algorithmic criteria. The

second one is representational bias, which means that the statistical imbalances in data collection will also influence the algorithm. This statistical violence disregards and excludes entire populations from consideration. Measurement bias uses standardized quantification to enforce specific modes of educational practices and thought, away from complex sociocultural realities.

2.2 Dual Obfuscation of Model Black Boxes: Erasure of Difference and Validation Failures

In this section we will explore two dual obscuration mechanisms that arise when deep learning models are used in education. The first is aggregation bias, which consists in erasing or smoothing away group particularities under the pretense of technical universality offered by the model. The second is evaluation bias, which emerges from the coloniality of the validation frameworks of these models. In both cases, deep learning models understand special educational needs as statistical outliers, rather than humanistic variables and introduces a homogenizing tool in the assessment systems.

2.3 Closed-Loop Traps in System Operation: Discriminatory Deployment and Bias Amplification

Self-reinforcing Discriminatory Ecosystem. The way that educational AI are designed to identify students who are “at risk” of dropping out creates self-reinforcing discriminatory ecosystems that lock in and institutionalize all manner of discrimination. The feedback loops activated by algorithmic biases create the Matthew effect with resource allocation. The initial “error” in algorithm design and deployment can easily result in resource deprivation that compounds statistical prejudice and perpetuates cycles of discrimination. The closed cycle of causation, while also digitizing and profiting discrimination in a sinister way, also actively engineers educational segregation by “personalized” learning path recommendations that lock students into ever-intensifying algorithmic discrimination circuits. The iterative upgrades of these systems are only deepening the damage and engineering a self-perpetuating architecture of inequality.

3 The Collaborative Governance Path to Deal with the Bias of Artificial Intelligence Algorithm in the Integration Construction of Ideological and Political Education Across Universities, Secondary, and Primary Schools

It is necessary to promote the integrated development of ideological and political education at all levels of education, and to adopt a collaborative governance approach involving technology, institutions, and other stakeholders to mitigate the bias effects of artificial intelligence algorithms. At the same time, it is necessary to embed the core values of ideological and political education into technological applications, so

that artificial intelligence and ideological and political education can achieve deep integration and mutual promotion.

3.1 Building a Precision-Oriented Data Governance System

One of the main sources of algorithmic ethical risks is thus the bias inherent to historical data. In the field of integrated ideological and political education for universities, secondary and primary schools, the educational big data currently in use suffers from sample bias and inappropriate feature selection, leading to implicit algorithmic biases in gender, behavior, and other dimensions. The majority of the work on algorithmic bias, including in education, has focused on the stages of the process where an algorithm is developed or evaluated. However, it is difficult for any algorithmic or metric-based approach to make a positive impact if appropriate data is not available to the algorithm. Therefore, it is imperative to advance data collection from simplistic “accuracy” to scientifically “precision,” establishing data quality evaluation standards and leveraging multidimensional verification, cross-validation, and other technical methods to correct historical data biases. Additionally, “intelligent cleaning tools should be developed to dynamically monitor the representativeness, completeness, and balance of datasets, ensuring input data objectively reflects real educational scenarios and blocking bias generation at its source”⁵.

3.2 Establishing a Systematic Regulatory Standards Framework

An ethical regulatory system covering the entire algorithm life cycle must be constructed. Key priorities include creating a tripartite evaluation framework centered on “transparency, explainability, correctibility,” adopting discrimination-aware data mining technologies to monitor algorithmic decisions in real time. Interdisciplinary regulatory teams should be formed to develop ethical review guidelines for AI products, clarify algorithm transparency grading standards, and require enterprises to regularly disclose sources of training data and decision logic. A risk early-warning mechanism should be established, implementing circuit-breaker protocols for biased applications to form a closed-loop regulatory process spanning “development-testing-deployment-iteration.”

3.3 Implementing a Multi-Stakeholder Collaborative Governance Mechanism

“A central overarching theme is the importance of situating the algorithm or AI being assessed within its particular socio-technical context—it is that socio-technical system and not merely the algorithm itself that is the proper object of assessment”⁶. A collaborative governance network involving developers, enterprises, schools, teachers, students, and parents must be built. Developers should embed ethical rule modules into algorithms, enterprises should institute beforehand evaluation systems for bias impacts, and schools should offer algorithmic literacy courses to cultivate digital citizenship awareness. Emphasis should be placed on dual empowerment through “technology + ideological education,” utilizing explainable AI technologies to help educa-

tors and students understand algorithmic decision logic. Direct user feedback channels should be established to enable stakeholders to participate in algorithm optimization. Concurrently, an industry ethics alliance should be formed to develop ethical guidelines for educational AI and promote self-regulatory conventions.

4 Peroration

Artificial intelligence algorithm bias threatens the integrated development of ideological and political education through the triple mechanism of data, model and system. It is manifested in the distortion of historical data, the resolution of model differences and the system closed-loop discrimination. Based on the Marxist view of science and technology, Three-Dimensional Governance Framework of “Institutional Supervision-Technical Correction-Subject Empowerment ”should be constructed: Dynamic Correction of Data Bias, Full-Cycle Algorithm Supervision, Special tools enhance transparency and realize the dynamic balance between tools and value rationality. In the future, we should promote the transformation of algorithm to “ education enabler ”, maintain ideological security, ensure the basic task of ideological and political education, and support the modernization of education.

5 Acknowledgment

The author would like thanks for the strong support of this research a major project of Chongqing education reform, “Achievement-driven, Collaborative Innovation, Research and Practice of integrated Construction System of Ideological and Political Courses in Universities, Primary and Secondary Schools”(Project No., 241027)

References

1. Murugan T ,P. K ,Abirami A ,“Driving Quality Education Through AI and Data Science,” IGI Global,pp. 26,Februray2025.(references)
2. Coenraad M . Making the invisible visible: Youth designs for teaching about technological and algorithmic bias [J]. International Journal of Child-Computer Interaction, 2024, 39 100634-<https://doi.org/10.1016/j.ijcci.2024.100634>.
3. L. Chen, P. Chen, and Z. Lin, "AI Education: A Review," IEEE Access, Volume 8, pp. 75264-75278,2020, doi:10.1109/ACCESS.2020.2988510.
4. Ifenthaler, D., Majumdar, R., Gorissen, P. et al. Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. Tech Know Learn 29, 1693–1710 (2024). <https://doi.org/10.1007/s10758-024-09747-0>
5. Baker, R.S., Hawn, A. Algorithmic Bias in Education. Int J Artif Intell Educ 32, 1052–1092 (2022). <https://doi.org/10.1007/s40593-021-00285-9>
6. Hasan, A., Brown, S., Davidovic, J. et al. Algorithmic Bias and Risk Assessments: Lessons from Practice. DISO 1, 14 (2022). <https://doi.org/10.1007/s44206-022-00017-z>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

