



Algorithmic Trust Bias Among College Students in the Age of Artificial Intelligence: A Study on Cognitive Generation Mechanisms and Value-Oriented Issues

Xiaohua Qiu, Weiwei Wang*

Central University of Finance and Economics, Beijing 100081, China

**Author to whom correspondence should be addressed.*

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Against the backdrop of artificial intelligence's deep integration into higher education, the stable feedback and efficient output of generative AI have led college students to develop algorithmic trust bias. This is manifested in overreliance on algorithmic outputs, diminished multi-source verification practices, and passive absorption of values. The underlying causes are associated with AI's technical advantages, reduced learning workload, and weakened critical judgment capabilities. Universities need to construct a governance system centered on multi-source input, conceptual reinforcement, critical thinking training, and value orientation. By regulating the use of AI and reconstructing cognitive chains, institutions can help students maintain cognitive subjectivity and mitigate structural risks in talent cultivation.

Keywords: Generative intelligence; Algorithmic trust bias; Value judgment; Higher education; Cognitive structure

Online publication: January 20, 2026

1. Practical generation and cognitive foundation of algorithmic trust bias among college students

1.1. Stable feedback from intelligent systems constitutes the initial inducement of trust bias

One of the most prominent features of generative intelligence is stable feedback, which provides an important starting point for the formation of algorithmic trust bias ^[1]. Firstly, the system can maintain a high degree of consistency in language structure, logical sequence, and presentation method when facing different tasks, enabling students to gain a strong sense of predictability in the process of use. Stable feedback significantly reduces uncertainty in the learning process, making students feel safe cognitively and emotionally, thereby being more willing to rely on the system to complete learning activities. Secondly, as the frequency of use increases, students will unconsciously mistake this stable experience for a symbol of knowledge reliability. Coherent language, complete paragraphs, and smooth logic easily make students have the illusion of accurate content, but they often fail to realize that the generation process relies on probability structures rather than genuine academic judgments. The familiarity effect gradually equates stable expression with reliable content, thereby transforming algorithms from tools into judgment references ^[2]. In addition, stable feedback makes students unwilling to face the inherent differences and conflicts in multi-source information. Knowledge in the real world

is ambiguous and contradictory, while generative content is often presented in a single path, making students feel that the burden of processing complex information is significantly reduced, thereby reducing the willingness to actively compare, understand, and verify^[3].

1.2. Reduced learning load prompts students to abandon multi-source verification

Generative intelligence has significantly reduced the learning load, leading to a continuous decrease in students' demand for multi-source verification. Firstly, learning tasks that previously required material searching, text reading, viewpoint collation, and cross-material comparison can now be completed with a single generation, compressing the learning chain into an extremely short time.

This high efficiency has led students to gradually abandon exploratory learning steps, prioritizing task completion over learning and understanding. Secondly, students weigh costs and benefits in task selection, and the complete framework provided by generative content makes them believe that multi-source reading is unnecessary. The motivation to search for materials declines, and patience for complex materials weakens, making multi-source verification gradually lose its significance. Students default that system outputs already have sufficient content density and logical structure, requiring no further verification^[4]. In addition, multi-source verification itself has high cognitive requirements, requiring students to compare different texts, identify differences, and form independent understanding amid conflicts. However, generative intelligence replaces diverse materials with unified expressions, depriving students of the opportunity to practice multi-source thinking and gradually forming cognitive habits of single-source input^[5]. Long-term dependence makes students lack comparative awareness and the multi-angle foundation required for forming judgments.

1.3. Instrumentalized learning methods promote the formation of algorithm-centered tendencies

The deep intervention of intelligent technology has led to an obvious tendency of instrumentalization in students' learning methods, thereby promoting the formation of algorithm-centered tendencies. Firstly, instrumentalized learning means that students regard learning as task completion rather than meaning understanding. Under this goal orientation, generative content has become the most suitable choice due to its high efficiency and neat structure, enabling students to obtain direct satisfaction at the level of assignment presentation^[6].

Secondly, when students view generated content as learning outcomes rather than analytical materials, system outputs begin to occupy a central position in the cognitive structure. They gradually take algorithmic results for granted as judgment bases rather than content that needs to be inspected or revised, transforming algorithms from auxiliary tools into cognitive centers. In addition, algorithm centralization is also reflected in the way knowledge structures are organized. Instead of constructing multi-dimensional frameworks through reading multi-source materials, students follow the linear structures of generated content for thinking. The unified logical sequences and paragraph patterns provided by algorithms are gradually solidified into students' default cognitive paths, compressing their subjective space and making it difficult for them to generate new viewpoints through critical, comparative, or dialectical methods^[7]. Finally, this transformation not only causes students to lose the ability to reflect on algorithmic content but also changes the starting point of thinking and the path of value absorption. Students no longer take the initiative to question the source, stance, or logic of content but naturally accept algorithmic structures, expanding algorithmic trust bias from the usage level to a deep cognitive model, ultimately affecting the formation of learning quality and academic capabilities.

2. In-depth impacts of algorithmic trust bias on cognitive structures

2.1. Single-source input causes cognitive structures to lose tension and tend towards single-path orientation

The formation of cognitive structures relies on the convergence of multi-source materials, the analysis of viewpoint differences, and the stimulation of theoretical conflicts. However, single-source input caused by algorithmic trust bias is

undermining this foundation. Firstly, as students increasingly rely on generative content, knowledge sources gradually shift from multi-text, multi-context, and multi-tradition to a single algorithmic model. The diversity of external knowledge supply is compressed, and cognitive input begins to present a one-way flow. Secondly, single-source materials cannot provide sufficient contrast and differences, leaving students lacking necessary tension stimulation in the learning process. Without the collision of multiple perspectives, cognitive activities struggle to form in-depth understanding and critical analysis, and knowledge structures gradually tend to be flattened. Thirdly, algorithmic content is usually generated based on high-frequency paths, and expression structures exhibit obvious linear progression characteristics.

After long-term exposure, students are likely to mistake this single path for the most natural cognitive method, thereby gradually solidifying it into a fixed thinking pattern and losing the ability to approach problems from different angles. In addition, single-source input also reduces students' sensitivity to complex problems, making it difficult for them to identify the ambiguity, contradiction, and structural tension behind problems, forming a tendency to avoid complexity^[8].

2.2. Weakened judgment capabilities prevent students from identifying viewpoint stances and theoretical differences

Judgment capabilities not only determine whether students can identify viewpoint differences but also whether they can maintain logical awareness and a clear stance in complex knowledge structures.

However, algorithmic trust bias weakens this key capability in multiple links. Firstly, after frequent exposure to generative content, students are likely to regard it as a stable and recognized form of knowledge expression, ignoring potential stance assumptions, theoretical bases, or historical contexts behind the content, resulting in a significant decline in stance sensitivity.

Secondly, algorithmic content is often presented in common expression patterns, failing to distinguish the argumentation strategies of different theories, making students gradually adapt to unified argumentation methods and become less sensitive in judging differences in methods. Thirdly, the templated structure of the system makes content on different topics present highly similar expression outlines. This structural similarity can make students mistakenly believe that there are no substantive differences between different arguments, thereby weakening their ability to identify theoretical differences. In addition, weakened judgment capabilities are also reflected in the ability to evaluate content. Students no longer take the initiative to consult original documents, verify material sources, or compare the depth, logic, and theoretical consistency of arguments, but directly accept the framework provided by algorithms, making judgment passive^[9].

2.3. Passive absorption of value stances creates tension between educational goals and technical logic

Algorithmic trust bias not only affects knowledge structures and judgment capabilities but also exerts a profound impact at the level of value systems, creating obvious tension between the value consciousness emphasized by higher education and technical logic. Firstly, when using generative content, students find it difficult to actively recognize the inherent value orientations, which often stem from corpus structures or mainstream models and are naturally absorbed through long-term exposure, resulting in students' lack of consciousness in value judgment. Secondly, algorithmic content has the characteristics of convergent expression, and the value dimension is often compressed into a limited framework in simplified expressions, reducing the sources of value stimulation for students and weakening their ability to perceive values. Thirdly, as students gradually rely on the unified expressions provided by the system, their value systems are prone to a shrinking trend. Students may understand the viewpoints generated by the model as universal stances, thereby ignoring other cultural, historical, or theoretical perspectives, making the value structure gradually thin^[10]. In addition, weakened value judgment capabilities make it impossible for students to maintain a clear stance when facing social issues, ethical problems, and public affairs, nor can they carry out in-depth value reflection—this is precisely the core goal of higher education in cultivating an independent spirit and rational personality^[11]. Finally, when technical logic, characterized by efficiency, conciseness, and structural unity, gradually replaces the depth, diversity, and value consciousness emphasized by educational logic, the direction of talent training is prone to deviation. The risk of educational goals being marginalized

by technical logic increases significantly, forming an unavoidable structural conflict ^[12].

3. Governance paths for colleges and universities to address algorithmic trust bias

3.1. Establish a multi-source knowledge supply system and restore cognitive flexibility

Multi-source input is the fundamental path to mitigate algorithmic trust bias and restore cognitive flexibility. Firstly, colleges and universities need to reconstruct original knowledge sources, strengthening classic literature, original theoretical texts, and high-quality academic works in the curriculum system to enable students to re-establish direct connections with knowledge traditions ^[13].

Original knowledge can provide the clearest theoretical boundaries and the most solid discourse sources, serving as the foundation for students' judgment capabilities. Secondly, colleges and universities should strengthen the input of heterogeneous materials, systematically introducing interdisciplinary content, historical materials, policy texts, and multicultural perspectives into classrooms and learning tasks. These materials can not only provide students with different knowledge logics but also effectively break the singularity of generative content, enabling students to maintain cognitive tension from multiple perspectives. Thirdly, colleges and universities should promote the construction of reading guidance courses, systematically training students' ability to select literature, screen materials, identify stances, and evaluate content, enabling them to distinguish key viewpoints and make rational judgments among a large amount of information.

Such courses can help students master reading methods, thereby improving their sensitivity to knowledge quality. In addition to enriching knowledge sources at the classroom teaching level, colleges and universities should use libraries, digital resource platforms, and open databases to build sustainable multi-source information access channels for students, enabling them to access extensive, high-quality knowledge materials both inside and outside the classroom. Finally, only on the basis of stable multi-source input can students' cognitive structures recover from the single-path tendency caused by algorithmic content to an open, diverse, and tense state, thereby forming an effective resistance to technical dependence.

3.2. Construct a judgment capability training system and reconstruct cognitive chains

The restoration of judgment capabilities relies on systematic, structured, and in-depth capability training, which must be carried out around the reconstruction of conceptual awareness, argumentation capabilities, and cognitive chains. Firstly, conceptual awareness is the starting point of judgment capabilities. Colleges and universities should offer courses on concept analysis, the history of theories, and guided reading of classic works, guiding students to understand conceptual boundaries, distinguish conceptual connotations, and clarify the position of concepts in different theoretical traditions. Clear conceptual awareness can prevent students from being guided by weak or vague concepts, enabling them to maintain theoretical clarity and accuracy in writing and discussions. Secondly, argumentation capability training is the core link in the formation of judgment capabilities. Colleges and universities can use case analysis, viewpoint comparison, theoretical conflict resolution, and close text reading to help students understand the logical basis and applicable scenarios of different argumentation methods. Only through continuous argumentation practice can students understand the complexity of argumentation and avoid inertial expressions brought about by the common structures of algorithmic outputs. Thirdly, the restoration of cognitive chains is crucial for eliminating algorithmic trust bias. Students need to experience a complete process in learning, including information collection, material collation, structural analysis, logical organization, and expression output, so that knowledge and judgment can be reconnected through practical operations. The more complete the cognitive chain, the more stable the judgment structure students can form, and the more limited the space for algorithms to replace thinking processes. In addition, colleges and universities should set up research-oriented tasks, practical writing, and interdisciplinary discussions to enable students to exercise their judgment capabilities in real problem environments, thereby extending classroom training to actual cognitive activities. Finally, through systematic training in concepts, argumentation, and cognitive chains, students can truly restore the depth and independence of judgment, preventing algorithmic trust bias from solidifying in cognitive structures.

3.3. Construct a value-oriented system and form educational norms in the intelligent era

The cultivation of value systems is a key link in resisting the impact of algorithmic trust bias, which is related to whether students can maintain value consciousness and clear stance in the intelligent era^[14]. Firstly, colleges and universities need to systematically integrate value education into professional courses, enabling students to understand the logic behind different value stances in specific contexts through theoretical analysis, historical cases, cultural comparisons, and discussions on real-world issues in the classroom, thereby improving their value judgment capabilities. Secondly, teachers should take the initiative to guide students to identify the stances behind content in the classroom, helping them understand that values are not abstract concepts but internal forces driving the formation of viewpoints by analyzing the source, target, and audience of discussions^[15]. Only on the basis of clarifying stances can students maintain value awareness in complex content. Thirdly, colleges and universities should promote discussion-based teaching, enabling students to analyze value differences through multi-angle exchanges, thereby forming a more mature value structure through ideological collision. Discussions can prompt students to shift from passive absorption to active construction, helping to avoid value shrinkage caused by algorithmic content. In addition, colleges and universities must establish norms for the use of intelligence for students, clearly explaining the boundaries of using generative systems in the learning, writing, and research processes. The purpose of the norms is not to restrict technology but to help students understand the scope of application of technology, ensuring that they do not regard technical results as value bases. Finally, value education, teaching guidance, and technical norms together constitute a systematic value protection mechanism, enabling students to maintain their dominant position in the intelligent era, identify value biases from algorithmic content, and maintain their own stances, ultimately avoiding the squeeze of technical logic on educational logic.

Funding

2024 Beijing Research Project on the Integration of Ideological and Political Education in Primary, Secondary and Tertiary Schools, “Research on the Integrated Teaching of Ideological and Political Courses in Primary, Secondary and Tertiary Schools from the Perspective of the ‘Three Adaptations’ Concept” (Project No.: XXSZ2024YB04); Theoretical Research Project on Party Building and Ideological and Political Work of Central University of Finance and Economics in 2025–2026, “Research on the Characteristics of College Students’ Network Social Circle Stratification and Discourse Guidance Strategies of Mainstream Ideology” (Project No.: DJD25005); the CUFU Postgraduate students support program for the integration of research and teaching, “Research on the Opportunities, Challenges and Paths of AI-Empowered Mainstream Ideological Construction” (Project No.: 2025216)

Disclosure statement

The author declares no conflict of interest.

References

- [1] Zheng XD, Hu KY, Yang JM, et al., 2025, Where Exactly Are Education and Learning Heading in the Age of Artificial Intelligence—An Examination Based on the Embodied Cognition Perspective. *Modern Distance Education Research*, 37(6): 11–23.
- [2] Li Y, Dai Y, 2026, Artificial Intelligence as a Method: Systematic Generation and Governance Reconstruction of the Integration of Educational Technology Talents. *Journal of Higher Education Evaluation*, 42(1): 102–112 + 185–186.
- [3] Shang ZC, Yan YH, 2023, The Educational Application of ChatGPT and Its Transformations and Ethical Challenges. *Journal of Northeast Normal University (Philosophy and Social Sciences Edition)*, (5): 44–54.

- [4] Liang JB, Xie Y, Tang ZH, 2025, Mechanism, Challenges and Countermeasures of Generative Artificial Intelligence Empowering the Digital Transformation of Education. *Modern Education Management*, (11): 15–23.
- [5] Yang DZ, He JJ, 2025, Philosophical Reflection on the Reform of School Science Education from the Perspective of Artificial Intelligence. *Journal of Dialectics of Nature*, (11): 100–107.
- [6] Ou JJ, 2025, Toward a “Harmonious Engineering”: Philosophical Evolution and Path Transformation of Engineering Education in the Age of Artificial Intelligence. *Chongqing Higher Education Research*, 14(1): 33–42.
- [7] Wei R, Duan SJ, 2025, The Form and Guidance Direction of AI Empowering the Dissemination of Ideological and Political Education Discourse in Colleges and Universities. *Educational Research*, (10): 124–134.
- [8] Anonymous, 2025, Artificial Intelligence Driving Educational Reform: Application Fields, Practical Challenges and Future Prospects. *Journal of the Chinese Society of Education*, (S1): 4–6.
- [9] Wang YM, Wang XY, Liu CC, 2024, Research on the Ethical Risk Management Framework for the Application of Generative Artificial Intelligence in Education. *E-education Research*, (10): 28–34.
- [10] Tong H, Yang YJ, 2024, Potential and Risk Research of Generative Artificial Intelligence in Educational Applications Based on the “Moralization of Technology” Theory. *E-education Research*, (7): 12–18.
- [11] Chai N, Cao BQ, 2025, From Value Alignment to Value Symbiosis: Ethical Risks and Governance of School Education in the Intelligent Era. *Modern Distance Education Research*, 37(6): 104–111.
- [12] Wang JH, 2025, Reconsidering the Relationship Between Education and Technology. *Open Education Research*, (5): 4–10.
- [13] Liang Y, Zhang W, 2025, The Theoretical Framework, Practical Paradoxes and Path Optimization of DeepSeek Empowering Ideological and Political Education in Vocational Colleges. *Inner Mongolia Social Sciences*, (6): 181–189.
- [14] Xuan RT, 2025, Research on the Application, Potential Risks and Countermeasures of AI Agents in Vocational Education. *China Computer & Communication*, (22): 205–211.
- [15] Su JJ, Ye HS, 2025, From Embodied Intelligence to Educational Innovation: Multi-Disciplinary Construction of Cognitive Growth Paths. *Open Education Research*, (5): 79–88.

Publisher’s note

ART AND TECHNOLOGY PRESS INC. remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.