From Student Profiling to Institutional Action: Governance Boundaries for AI-enabled Student Support in Chinese Universities

Hongda Jiang[1]

[1] *School of Marxism, Nanjing University of Finance and Economics, Nanjing, China*
[1]*Corresponding author. Email: 9119951045@nufe.edu.cn*

ABSTRACT

AI-enabled student-support systems increasingly translate digital traces into profiles, risk estimates, and recommendations for educators. This study asks how a provisional description becomes actionable in Chinese universities. Through conceptual reconstruction, normative interpretation, and structured mapping of official documents from four universities, it develops a five-transition framework: data proxying, profile labeling, probabilistic inference, educational intervention, and feedback. Public documents identify functional entry points but do not verify effects or continuous mechanisms. The paper argues that digital stereotyping arises only when a purpose, evidence, action, or time boundary is crossed, institutional treatment changes, and correction mechanisms fail. It then specifies input, decision, action, and exit controls that keep high-impact transitions explainable, contestable, suspendable, and correctable. The framework connects human-centred learning analytics with the institutional use of student profiles while explicitly limiting empirical claims.

Keywords: *AI-enabled student support, Student profiling, Actionable labels, Learning analytics ethics, Digital stereotyping.*

1. INTRODUCTION: FROM TARGETED IDENTIFICATION TO INSTITUTIONAL JUDGMENT

China's "AI Plus Education" Action Plan calls for intelligent agents in ideological-political education and student development. Official materials from Beijing University of Posts and Telecommunications (BUPT), Harbin Institute of Technology (HIT), Tsinghua University, and Huazhong University of Science and Technology (HUST) describe data integration, dynamic profiles, targeted identification, growth archives, early warning, and counselor support.[1][2][3][4][5] These materials establish functional entry points, not platform effects. Course activity, consumption changes, movement patterns, and interaction texts remain contextual proxies. Once displayed as scores, alerts, or recommendations, however, they may acquire institutional actionability beyond their evidential warrant. The question is therefore not simply whether a profile is accurate, but through which transitions it shapes educational treatment and under what conditions targeted support becomes digital stereotyping.

Four strands of scholarship frame the problem. The first, on precision ideological and political education and student profiling, runs from object segmentation and demand matching to data-driven approaches, explaining how ideological and political work shifts from general supply to tailored intervention.[6][7] Huang Wenlin (2021) uses data resources, hierarchical labels, and profile models for personalized analysis and academic early warning.[8] Duan Hongtao (2026) builds an implementation system of "precise classification—precise content—precise method—precise evaluation" around college students' online behavior profiles.[9] Mi Huaquan and Wu Hongni (2026) discuss the operating mechanisms, dilemmas, and governance approaches of AI agents empowering precision ideological and political education.[35] What remains unaddressed is how profile labels enter organizational workflow, trigger actions of varying intensity, and how action records feed back to extend the validity of original judgments.

The second tradition, on intelligent ideological and political education, recommendation algorithms, and generative AI, has moved past a unidirectional empowerment narrative. Hu Hua (2022) reveals tensions between technological efficacy, rule logic, and the humanistic character of ideological and political education.[10] Mi Huaquan (2023), Huang Shihu and Wang Xunuo (2025) discuss the ethical risks of intelligent ideological and political education and how recommendation algorithms reshape educational narrative.[11][12] Generative AI research points to the reconstruction of subject-object relations and teaching affordances.[13][14] Yet "what content the agent generates for students" differs from "what judgment the agent generates for educators based on student data." The second path involves the interface of technical inference, organizational permission, and professional action. Its risk is not only content truthfulness but how system recommendations change the real pedagogical relationship.

The third tradition, on educational data ethics, has discussed subject-object dislocation, data leakage, algorithmic risk, and educational injustice arising from digital mapping.[15] International learning analytics research shows that students recognize the potential value of data analysis and at the same time draw boundaries on what data go to which actors.[25] Privacy risks and a sense of control affect students' trust in higher education institutions and their willingness to disclose information.[26] Specifying data collection practices can raise students' awareness of what is collected,[27] but being "informed" does not mean knowing how labels are generated, who sees them, what action they trigger, or whether corrections synchronize downstream. Rights principles have to land at the concrete interface from data to action.

The fourth tradition, on predictive learning analytics, has pushed the question from models to interventions, organizational workflow, and algorithmic fairness. Rets, Herodotou, and Gillespie (2023) argue that ethical requirements must be embedded in institutional policy, tool development, and actual workflow.[24] Dai et al. (2025) show that risk prediction alone does not improve learning, and feedback intervention is an independent step.[28] Kizilcec and Lee (2022) and McConvey and Guha (2024) document that different groups can bear different false-positive, false-negative, and resource consequences.[29][30] Froehlich and Weydner-Volkmann (2024) show that adaptive interventions can be designed through ethical reflection and fairness considerations to mitigate identity threat and promote equity.[31] Two propositions therefore do not hold: that algorithmic labels necessarily harm students, and that more accurate models make intervention more legitimate.

This study treats label enactment as the process through which a digital label becomes actionable through authorization, interface presentation, workflow invocation, and consequential use. Conceptual reconstruction distinguishes profile, enactment, and stereotyping; normative interpretation separates legal duties, standards, ethical principles, and the author's proposals; structured mapping records whether the four universities' public documents disclose data, profile, output, and action interfaces. Only official documents verified by title, date, and issuing body were admitted by August 13, 2026. "Not disclosed" means no evidence in the admitted materials, not institutional absence. Academic warning, psychological support, or administration enters the analysis only when outputs feed into value guidance, educational relationships, focused attention, or resource allocation. Without internal logs, complete institutional rules, interviews, or case records, the study makes no claim about the frequency, causal strength, or actual risk level of any platform.

2. DIGITAL PROFILES ARE NOT INSTITUTIONAL JUDGMENTS: CONCEPTUAL BOUNDARY AND UNIT OF ANALYSIS

A "student growth agent" is not a student profile or a general chatbot. In Chinese policy and university practice, the term refers to an AI-enabled student-support system that connects longitudinal student data, model-based inferences, institutional permissions, and educator workflows. This study treats it as a digital system that, over a sustained period, continuously pulls student data and interaction information, generates inferences and recommendations from models or rules, and connects with higher education institutional systems, organizational permissions, and educator behavior. Sustainability, inferentiality, and organizational coupling are three features that set it apart from general information platforms.

This distinction matters more in digital ideological-political education and student support in Chinese higher education than in most other settings. A misjudged e-commerce recommendation usually means a piece of content goes unclicked. A misjudged pedagogical output can be read by actors with organizational authority and professional legitimacy and feed into teacher-student relationship, professional support, and on-campus management. The same “low activity” label, used to deliver a closable resource reminder to the student, is not the same as queuing the student onto a focused conversation list. The first path carries risk mainly of low relevance and information disturbance. The second path adds evidence, permission, human judgment, and student statement. The unit of analysis is therefore not simply whether a label is accurate, but what action it triggers under which authorization, interface, and procedure.

A digital profile is a structured representation built for a particular purpose, data scope, and classification rule. It is a selective description of certain recordable student characteristics, not the student as a person. A label is the categorical unit within the profile that can be retrieved, calculated, compared, or invoked. Bowker and Star (1999) show that categories are not neutral placements of reality into cells; they distribute visibility, resources, and treatment within organizations. [20] Becker (1963) shows that categories acquire real effect only through the actual application of social rules and sanctions.[21] The interactive variation between the classified and the classification is carried by Hacking’s looping effects, discussed later. Labels do not always harm. Under specific institutional conditions, positive academic labels can also lift students’ self-concept and development expectations.[22]

Pedagogical judgment is the professional and organizational judgment made by educational actors in concrete scenarios about students’ needs, risks, modes of support, or developmental possibilities. It can decide who is attended to, which resources are obtained, what conversation or referral occurs. The “label enactment” used in this study is a process concept with three layers. The first is the precondition: the label is visible to actors with organizational authority or institutional systems and connects with action options such as resource recommendation, attention ordering, conversation, referral, early warning, or evaluation recommendation. The second is the process event: the label is actually invoked in specific institutional practice and becomes input for ordering, prompting, recommendation, review, or action. The third is the consequence: whether invocation changes the attention, resources, relationships, or opportunities the student receives, and for how long. The first two layers together decide whether enactment occurs. The third layer evaluates intensity and consequence, and does not let the outcome redefine the process backwards.

Counterexamples draw the boundary. A self-management label shown only to the student and not entering any downstream workflow is “labeled but not enacted.” A label connected to action interfaces but never invoked in specific institutional practice has only the condition, not the event, of enactment. Providing closable resources based on the profile with real-time correction is “enacted but not stereotyping.” Teachers’ judgments based entirely on face-to-face experience can produce similar label effects but do not count as digital label enactment. Even when a digital label triggers conversation, enactment need not turn into stereotyping as long as educators can refuse, student statements can alter the judgment, and labels expire on schedule. Label enactment is the process from representation to action. Digital stereotyping is the possible risk outcome of this process under conditions of boundary crossing and error-correction failure.

3. FROM DESCRIPTION TO ACTION: A FIVE-TRANSITION CHAIN OF LABEL ENACTMENT

Label enactment can be broken into five transitions. Student state is translated into data proxy. Data proxy is organized into profile labels. Profile labels are converted into probabilistic inference. Probabilistic inference enters pedagogical intervention. Intervention results loop back as new data and labels. This is a theoretical decomposition of the institutional workflow. It does not assume that every platform runs all five steps or that any step necessarily leads to the next.

The first transition is from student state to data proxy. The agent does not directly observe students’ thoughts, emotions, or situations. It builds a proxy relationship through recordable data such as access control, consumption, login, course performance, activity participation, and interaction text. The first choice of what data to collect already decides what is visible and what stays outside the system. An absence may point to academic difficulties, or it may be caused by illness, family circumstances, internship, or systematic record bias. A proxy can open understanding, but it cannot by itself become proof of internal state.

The second transition is from data proxy to profile labels. Data go through cleaning, aggregation, label definition, and weight setting to form categories such as poverty risk, academic fluctuation, declining activity, or focused attention. This step is not the algorithm's simple discovery of objective facts. It is the organization's joint decision on which differences deserve naming, what standards set the boundary, and who will see the results. The more the label vocabulary resembles personality or identity judgment, the easier it is for the label to expand across scenarios, and the easier it becomes to lose the difference between the profile and the student as a person.

The third transition is from profile labels to probabilistic inference. Profiles can be turned by rules or models into risk scores, similar groups, trend predictions, or action recommendations. Two leaps are easy to overlook: from group correlation to individual causation, and from probability to identity. Average accuracy does not answer whether this particular student warrants action, and it does not show whether false positives and false negatives concentrate in particular groups. High-impact outputs therefore have to keep "to be verified," "insufficient information," and alternative explanations visible, and should not compress probability into a single color while granting it certainty.

The fourth transition is from probabilistic inference to pedagogical intervention. Inference changes educators' attention ordering and problem framing through list ordering, risk colors, message reminders, or recommendation scripts. The presence of humans in the workflow does not automatically mean substantive human judgment. If reviewers see only conclusions without access to data sources and uncertainty, and if accepting recommendations is the default while refusal requires extra justification, human review can degenerate into a shell of accountability. Du, Liu, and Xian (2026) ran a 2×2 controlled experiment with 214 university teachers using fictitious student essays and self-reported intervention intentions, and found that, with the same text under evaluation, AI detection alerts and visual risk prompts shifted some teachers' evaluation and intervention judgments.[32] This result does not directly prove that student profiles produce the same effect, but it does indicate that algorithmic prompts can shape the judgment environment.

The fifth transition is from intervention results to feedback loop. Students' avoidance, anxiety, or low participation after being focused upon can be re-recorded and read as confirmation of the original risk label. Educators' increased conversation records prompted by the system can serve as evidence of the student's continuing need for attention. Hacking (1995) shows that the classified will respond to classification and the treatment it produces, and that this response changes the object to which the category refers.[33] Feedback can also correct rather than fixate labels. When students can supply context, educators can record counterevidence, the system can distinguish pre- and post-intervention data, and labels expire on schedule, the loop can shift from self-confirmation to joint correction.

The five transitions are not a causal model verified through statistical testing. They are an analytical chain that can be confirmed, revised, or refuted by process documents, logs, interviews, and case materials. Any transition can be suspended. Proxy signals can be marked with limitations. Labels can be restricted in use. Probability can remain a hypothesis to be verified. Educators can refuse system recommendations. Students can cause old labels to expire through statements and correction. The point of the analytical chain is to mark the interfaces from representation to action, not to claim that technology necessarily leads to stereotyping.

Institutional actionability can be amplified semantically when behavioral traces receive identity-like names, visually when uncertainty becomes ranks or colors, and organizationally when labels are widely visible and tied to workflows. These are candidate conditions, not verified mechanisms. Two propositions remain testable: stronger semantic, interface, and organizational amplification should increase direct label adoption and reduce human modification; effective student correction, counterevidence, downstream synchronization, and expiry should shorten erroneous judgments. Process documents, logs, interviews, interface experiments, refusal records, and expiry data could confirm or refute these propositions.

4. PUBLIC-DOCUMENT MAPPING OF FOUR UNIVERSITIES: FUNCTIONAL ENTRY POINTS AND GOVERNANCE INFORMATION GAPS

Beijing University of Posts and Telecommunications, Harbin Institute of Technology, Tsinghua University, and Huazhong University of Science and Technology are included not as a representative sample of Chinese universities. They are included because their official documents disclose, with relative concentration, functions of data, profiling, identification, early warning, and organizational coordination. Materials were structurally mapped along data sources, profile content, system outputs, manual review, student rights, responsible actors, and traceability information only. The mapping did not compare case effects or verify the transition analysis chain. “Not disclosed” means that admitted documents provide no evidence, not that no corresponding institution exists within the university.

BUPT: Official materials describe cross-departmental data integration, student growth profiles, alerts, resources, and action recommendations entering student-affairs work. They also mention data governance and project compliance evaluation. Detailed procedures for individual label explanation, contestation, and expiry are not disclosed. The permissible conclusion is limited to the existence of connections among data, profiling, and action interfaces.[2]

HIT: Official materials describe multidimensional data awareness and a “student–agent–counselor” interaction structure supporting ideological-political work. Collaboration among student-affairs and information-technology departments is disclosed, whereas model thresholds, student correction, and label duration are not described in detail. The materials establish conditions for continuous perception and human–machine collaboration, not their effects.[3]

Tsinghua University: Official materials connect classroom and extracurricular data with growth support, intelligent assistants, identification of students requiring targeted support, and service provision. University-level coordination is disclosed, but evidence thresholds, review records, and contestation interfaces for individual identification are not described in detail. The materials show a functional gradient from service recommendation to higher-impact identification.[4]

HUST: Official materials describe growth archives, knowledge support, academic warning, crisis prediction, and counselor decision support. Multi-layer knowledge construction and risk assessment are mentioned, while case-level permissions, contestation, revocation, and deletion records are not described in detail. The materials establish conditions for connecting archives, prediction, and professional action.[5]

Evidence for this mapping comes from BUPT, HIT, and HUST materials released by the Ministry of Education’s Department of Ideological and Political Work on June 9, 2025, and Tsinghua materials released on December 18, 2025.[2][3][4][5] Horizontal mapping first reveals a functional gradient from information services to high-impact early warning. Low-impact functions mainly provide general Q&A, resource navigation, and closable reminders. Medium-impact functions begin providing growth suggestions, mentor recommendations, or proactive conversation reminders based on individual data. High-impact functions may involve crisis early warning, identification of students requiring targeted support, or recommendations affecting opportunities. All four universities’ public documents contain functional entry points of different intensities. That fact demonstrates that graduated evidence and procedures are practically necessary. It does not demonstrate that these functions have produced equivalent consequences.

Functional disclosure and governance process disclosure are asymmetric. Public materials clearly explain what the system can do and how organizational collaboration is arranged. They are usually insufficient to answer three branches. First, whether the output is a low-impact reminder, professional advice, or a final action that already changes the student’s treatment. Second, whether students can see, supplement, correct, or raise dissent against high-impact outputs. Third, how the system distinguishes pre- and post-intervention data so that old labels expire, are revoked, and synchronize downstream. This disclosure gap should be treated as a governance issue in its own right, but it cannot be used to infer specific universities’ violations, harm, or absence of internal institutions.

The four universities’ public documents at most support L1-level structural risk research. Multi-source data, profile inference, and action interface are already juxtaposed in the institutional narrative, requiring scrutiny of the transitions between them. To establish L2-level boundary-crossing intervention, actual label definitions, model thresholds, case evidence, permissions, and action records are required. To establish L3-level digital stereotyping, it must be further shown that boundary-crossing judgments persistently change student treatment and that dissent, correction, review, or label expiry fail to suspend the consequence. The current public evidence can verify neither that the five transitions occur continuously on specific platforms nor that L2 or L3 determinations are warranted for the four universities.

5. RISK DETERMINATION: WHEN DOES TARGETED SUPPORT DEGRADE INTO DIGITAL STEREOTYPING

Treating all data collection, label generation, or risk prompting as digital stereotyping would make the concept indistinguishable. It would also erase the potential value of data tools in discovering overlooked needs and expanding support coverage. This study proposes three minimum conditions. First, at least one of the four boundaries (purpose, evidence, action, or time) is substantially crossed. Second, the boundary-crossing judgment enters organizational processes and changes how students are attended to, receive resources, accept conversations or referrals, and participate in evaluation. Third, mechanisms of explanation, review, correction, dissent, or label expiry are insufficient such that erroneous or outdated judgments continue to exert influence. Targeted support degrades into digital stereotyping only when “boundary crossing, action consequence, and error-correction failure” overlap.

An evidence gradient must match claim strength. Official functional descriptions can establish only L1 structural risk. L2 requires case-specific proof of a boundary-crossing intervention, including the label definition, relevant data, threshold, permissions, human review, and action record. L3 additionally requires a documented persistent consequence and failure of contestation, correction, synchronization, or expiry. Limited disclosure and possible model error alone cannot support L2 or L3.

Purpose crossing: Data existence does not automatically confer legitimacy for entry into any pedagogical scenario. Data use has to justify its direct relevance to educational tasks, less-impactful alternatives, and reassessment after use changes. The Personal Information Protection Law requires purposes to be specific, reasonable, and directly related, and the use of the least-impactful means. [16] Educational data classification and grading standards give schools a reference for data governance. [17] Purpose crossing shows up mainly as function drift: academic support data entering qualification evaluation, or psychological support interaction being spliced into general management profiles. The distinction is not whether schools can accommodate management. It is whether students obtaining appropriate help remains a verifiable goal, or whether improving controllability, reducing cost, or evading responsibility becomes the only substantive outcome.

Evidence crossing: Evidence crossing includes the leap from behavioral trace to internal state, from group correlation to individual causation, and from risk probability to stable identity. Algorithmic fairness research shows that bias can enter measurement, labels, models, and action stages, causing particular groups to bear more false-positive intervention or false-negative loss. [29][30] Evidence thresholds have to match output intensity. General recommendations can tolerate lower certainty. Crisis early warning, identification of students requiring targeted support, and opportunity-affecting recommendations cannot apply the same threshold. Proxy variables can serve as clues for further understanding, but outputs must keep “to be verified,” “uncertain,” and alternative explanations visible, and must absorb students’ contextual statements.

Action crossing: Even without automatically making final decisions, systems can change pedagogical action through attention ordering, conversation lists, and problem framing. The Personal Information Protection Law requires automated decisions to be transparent, fair, and impartial, and grants explanation and the right to refuse decisions made solely by automated decision-making where the impact is significant. [16] Whether specific on-campus interventions meet this legal threshold requires case-by-case judgment. When systems approach important opportunity or rights-and-interests decisions, a formal “someone in the back office” cannot substitute for comprehensible explanation and genuine human judgment. The analytical scope of label enactment is broader than the legal sense of automated decision-making. Even when the final decision is nominally made by a person, as long as digital labels have been invoked and altered the judgment process, the case can enter this analysis. That does not mean all enactment automatically triggers the same rights consequences as Article 24 of the Personal Information Protection Law. Determination still has to combine whether it constitutes automated decision-making, the degree of impact, and case facts.

Action control has to prevent both erroneous and missed interventions at once. Low-impact recommendations can be automatically executed with feedback. Medium-impact suggestions require educator verification and student contextual input. High-impact alerts require professional review, reason recording, and dissent channels. The substance of crossing is the loss of proportion between evidence strength, human judgment, and real-world consequence.

Time crossing: Stage-specific difficulties can trigger support at the time but should not indefinitely explain subsequent behavior. Time crossing shows up as labels without expiry, scenario-specific labels migrating across domains, or intervention records inversely confirming the original label. The principles of shortest necessary retention, correction, and deletion in personal information protection do not directly stipulate the period for each type of educational record, but they negate “once useful, permanently valid.” [16] Time governance has to distinguish retention of original facts, visibility of inferred labels, effect of old judgments, and audit use for error correction. When students supply context, educators record counterevidence, and labels expire on schedule and synchronize downstream, feedback shifts toward joint correction. Accumulating only data supporting the original judgment can fixate stage difficulties into long-term identity.

6. MINIMUM CONTROL RESPONSIBILITY: KEEPING INSTITUTIONAL TRANSITIONS SUSPENDABLE

Stating that “schools, enterprises, teachers, and students jointly govern” by itself may diffuse rather than concentrate responsibility. Gong and Yan (2026) have classified educational algorithms into passive-execution, guided-feedback, and deep-intervention types according to the depth of algorithmic participation and discussed different legal attribution schemes for each.[36] “Stratified responsibility” therefore is not the innovation of this study. The study does not presume administrative, civil, or other legal liability in the absence of case facts. It addresses a narrower question: in each transition of label enactment, who must at minimum retain the capacity for explanation, refusal, suspension, or repair.

Minimum control responsibility rests on three capacities. Actors who can decide whether a transition enters the institutional workflow bear the control obligation. Actors who hold data, models, or student context bear the explanation and verification obligation. Actors who can revoke erroneous labels and stop downstream influence bear the repair obligation. This is not about evenly distributing responsibility to all participants. It is about anchoring responsibility at the interface most capable of suspending risk. The Personal Information Protection Law provides purpose limitation, permission control, impact assessment, and rights response bases.[16] JY/T 0643-2025 and JY/T 0661-2025 set standards for personal information protection in smart education platforms and for educational data classification and grading.[17][34] AI ethics specifications and learning analytics ethics research provide lifecycle and workflow embedding principles.[18][23][24] The remaining content of the table below belongs to the minimum governance and professional requirements proposed by this study on this basis, and does not equate to direct legal liability determination. The actor names in the table indicate only that corresponding governance functions must have identifiable bearers, without presuming each school’s specific institutional settings. “Primary function bearer” indicates who should configure and preserve control capacity in the corresponding transition, not exclusive principal responsibility, ultimate legal liability, or exemption from responsibility for other participants.

This framework can be further compressed into four control points. Input control answers “why identification is permitted.” It requires that a specific pedagogical scenario precede the decision about which data to integrate, not the other way around. Decision control answers “on what basis judgments are formed and used.” It requires that inferential labels simultaneously present data timeliness, main basis, uncertainty, alternative explanations, applicable scenarios, and expiry. The higher the action impact, the higher the requirements for evidence quality, human review, and reason recording. Action control answers “who may take what measures accordingly.” It requires low-impact resource recommendations to be closable with feedback, medium-impact suggestions to be verified by educators with student statement input, and high-impact alerts or opportunity recommendations to be subject to professional review and dissent examination. Exit control answers “when does judgment expire.” It requires labels to stop triggering actions upon expiry, corrections to synchronize downstream, and reactivation to require new data and new judgment.

Student participation is not an unconditional veto of all educational management, and students should not have to share responsibility for model bias. Students’ contextual explanations of absence, consumption changes, or low activity are important evidence for testing proxy variables and probabilistic inferences. Educators and professionals can only assume substantive judgment responsibility when they can access evidence, understand limitations, and genuinely refuse system recommendations. Technical actors must ensure label explanation, model version, interface uncertainty, revocation synchronization, and minimum-necessary logs are achievable, but should not substitute for authorized educators’ decisions on specific actions. Higher education institutions, as the organizational authorizers for systems entering educational workflows, must at minimum ensure these capacities genuinely exist, and cannot use outsourcing or back-end human signature to substitute for scenario review.

Emergency life-and-health scenarios may permit minimum, necessary, and reversible protective measures. But emergency exceptions must have triggering conditions, action limits, and post-hoc review. They cannot be turned into unbounded authorization for continuous monitoring. Logs and audit are also subject to minimum-necessary and limited-retention constraints. The Measures for Science and Technology Ethics Review (Trial) can serve as reference for scientific research and technology development activities, but cannot, without applicable analysis, directly become a statutory procedure for all on-campus daily applications.[19]

Minimum control responsibility has to be calibrated by care outcomes. Model recall rate, alert count, and conversation count only indicate system activity, not pedagogical effectiveness. Limited pilots should observe at the same time whether genuine support arrives, whether unnecessary interventions decrease, whether human modifications and student dissent alter results, how long erroneous label withdrawal takes, and whether false positives, false negatives, and resource access show unexplained group differences. Only when “more identification” passes the three tests of intervention proportionality, error-correction effectiveness, and student subjectivity can it be interpreted as “more targeted support.”

7. CONCLUSION: THE LIMITS OF INSTITUTIONAL JUDGMENT IN AI-ENABLED STUDENT SUPPORT

Student growth agents have brought digital ideological-political education and student support in Chinese higher education into a stage of continuous perception, dynamic profiling, probabilistic inference, and action recommendations linked together. The question is no longer whether profiles can support classification. It is under what conditions finite digital descriptions acquire organizational actionability and change the attention, support, and opportunities students receive.

“Label enactment” offers a meso-level process concept in response. Organizational visibility and action interface constitute the precondition. The actual invocation of a label in specific institutional practice constitutes the enactment event. Whether and for how long student treatment changes is an outcome that requires separate observation. The conceptual contribution, compared with classification consequences, label imposition, and the algorithmic decision chain, is to identify permissions, interfaces, invocations, and institutional workflows as observable interfaces through which digital representation enters organizational action. Labels not invoked do not constitute enactment. enactment subject to genuine refusal, student statement, and label expiry does not necessarily form digital stereotyping.

Data proxying, profile labels, probabilistic inference, pedagogical intervention, and feedback loop form a transition analysis chain awaiting empirical testing, not a causal mechanism already verified by the four universities’ materials. The four universities’ official public documents only show that functional entry points of data, profiling, early warning, and educator support are already juxtaposed in real institutional narratives, enough to support L1-level review necessity. They cannot prove that the five transitions occur continuously on specific platforms, nor can they support determinations of boundary-crossing intervention or digital stereotyping.

Label enactment is not equivalent to digital stereotyping. Targeted support degrades into digital stereotyping only when purpose, evidence, action, or time boundaries are substantially crossed, boundary-crossing judgments enter organizational action, and explanation, correction, dissent, or label expiry fails to suspend the consequence. Responsibility should not be expanded into multiple parallel governance lists but should rest on four minimum control points: who can release data and scenarios, who must explain labels and inferences, who can refuse disproportionate intervention, and who can revoke old labels and stop downstream effects. Legal obligations, industry standards, ethical principles, and professional requirements differ in normative strength. All of them should be turned into genuinely existing suspension and repair capacities.

The four universities' materials support only L1 structural review, not findings that a continuous mechanism, boundary-crossing intervention, or digital stereotyping actually occurred. Future studies should test the chain through permissions, invocation logs, human modification, student dissent, correction synchronization, expiry, and consequence records. This normative-theoretical study therefore does not propose eliminating uncertainty through more data. It argues that profiles may open an inquiry and mobilize care, but must remain provisional judgments that can be explained, refused, corrected, suspended, and ended.

ACKNOWLEDGMENTS

This paper is a phased outcome of the 2023 Jiangsu Provincial Higher Education Philosophy and Social Sciences Research Fund Project "Research on Xi Jinping's View of Network Ideological and Political Education" (Project No. 2023SJSZ0118) and the 2023 Teaching Reform Project of Nanjing University of Finance and Economics "Research on AI Technology Empowering the High-Quality Development of Ideological and Political Education in Universities."

REFERENCES

[1] Ministry of Education and Four Other Departments. "AI Plus Education" Action Plan. (2026-04-02) [2026-08-13]. https://www.moe.gov.cn/srcsite/A16/s3342/202604/t20260410_1433240.html.

[2] Department of Ideological and Political Work, Ministry of Education. Beijing University of Posts and Telecommunications Focusing on "Four-Dimensional Collaboration" to Build a New Digital-Intelligent Ideological and Political Education Landscape Relying on the Student Growth Agent Platform. (2025-06-09) [2026-08-13]. https://www.moe.gov.cn/s78/A12/gongzuo/moe_2154/202506/t20250609_1193458.html.

[3] Department of Ideological and Political Work, Ministry of Education. Exploration and Practice of Ideological and Political Education Empowered by Intelligent Agents at Harbin Institute of Technology. (2025-06-09) [2026-08-13]. https://www.moe.gov.cn/s78/A12/gongzuo/moe_2154/202506/t20250609_1193460.html.

[4] Department of Ideological and Political Work, Ministry of Education. Tsinghua University Actively Advancing AI-Empowered Ideological and Political Education. (2025-12-18) [2026-08-13]. https://www.moe.gov.cn/s78/A12/gongzuo/moe_2154/202512/t20251218_1423781.html.

[5] Department of Ideological and Political Work, Ministry of Education. Huazhong University of Science and Technology Building a New Paradigm of AI-Empowered "Three-Way Interaction" Ideological and Political Education. (2025-06-09) [2026-08-13]. https://www.moe.gov.cn/s78/A12/gongzuo/moe_2154/202506/t20250609_1193457.html.

[6] ZHOU Y. Precision Ideological and Political Education: A New Concept and Model for Ideological and Political Work in Universities in the New Era. Ideological and Theoretical Education, 2020(8): 100-105.

[7] DING K, SONG L Z. On the Mechanism and Realization Path of Precision in Ideological and Political Education in Universities. Ideological and Theoretical Education, 2020(6): 101-105.

[8] HUANG W L. An Exploration of Precision Ideological and Political Education in Universities Based on Student Profile Analysis. Journal of Northeastern University (Social Science), 2021, 23(3): 104-111. doi:10.15936/j.cnki.1008-3758.2021.03.014.

[9] DUAN H T. Logical Approach and Implementation Strategy for Precision Ideological and Political Education Based on College Students' Online Behavior Profiling.

Ideological and Theoretical Education, 2026(1): 90-97. doi:10.16075/j.cnki.cn31-1220/g4.2026.01.013.

[10] HU H. Intelligent Ideological and Political Education: The Integration of Ideological and Political Education and Artificial Intelligence in the Era. Ideological Education Research, 2022(1): 41-46.

[11] MI H Q. Generation Logic, Manifestations, and Prevention-Control Mechanisms of Ethical Risks in Intelligent Ideological and Political Education. China Educational Technology, 2023(2): 111-117.

[12] HUANG S H, WANG X N. Introduction, Alienation, and Control: Risk and Response of Ideological and Political Education Narrative Based on Recommendation Algorithms. Journal of Hohai University (Philosophy and Social Sciences), 2025, 27(1): 9-17. doi:10.3876/j.issn.1671-4970.2025.01.002.

[13] LI C M, ZHANG F. Reshaping the Subject-Object Relationship in Ideological and Political Education through Generative Artificial Intelligence. Journal of Jishou University (Social Sciences), 2025, 46(4): 103-110. doi:10.13438/j.cnki.jdxb.2025.04.012.

[14] ZHAO W, DOU X R. Affordance and Constraint Analysis of Generative AI Empowering the Teaching Reform of Ideological and Political Courses in Universities. Journal of University of Electronic Science and Technology of China (Social Sciences Edition), 2026, 28(1): 148-157. doi:10.14071/j.1008-8105(2025)-3043.

[15] DIAO S F, KONG X Y. Ethical Issues in the Application of Digital Twin Technology in Education. Journal of South China University of Technology (Social Science Edition), 2022, 24(6): 16-24. doi:10.19366/j.cnki.1009-055X.2022.06.003.

[16] Standing Committee of the National People's Congress. Personal Information Protection Law of the People's Republic of China. (2021-08-20) [2026-08-13]. https://www.npc.gov.cn/npc/c2/c30834/202108/t20210820_313088.html.

[17] Ministry of Education. Guidelines for Educational Data Classification and Grading: JY/T 0661-2025. (2025-12-18) [2026-08-13]. https://www.moe.gov.cn/srcsite/A16/s3342/202601/W020260107510856466340.pdf.

[18] National New Generation Artificial Intelligence Governance Professional Committee. Ethical Norms for New Generation Artificial Intelligence. (2021-09-25) [2026-08-13]. https://www.most.gov.cn/kjbgz/202109/t20210926_177063.html.

[19] Ministry of Science and Technology and Ten Other Departments. Measures for Science and Technology Ethics Review (Trial). (2023-09-07) [2026-08-13]. https://www.most.gov.cn/xxgk/xinxifenlei/fdzdgknr/fgzc/gfxwj/gfxwj2023/202310/t20231008_188309.html.

[20] BOWKER G C, STAR S L. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press, 1999: 5-6.

[21] BECKER H S. Outsiders: Studies in the Sociology of Deviance. New York: Free Press, 1963: 9.

[22] FENG Q, WANG X. The Psychological Effects of Academic Labeling: The Case of Class Tracks. Journal of Comparative Economics, 2018, 46(2): 568-581. doi:10.1016/j.jce.2017.10.004.

[23] BUCKINGHAM SHUM S, FERGUSON R, MARTINEZ-MALDONADO R. Human-Centred Learning Analytics. Journal of Learning Analytics, 2019, 6(2): 1-9. doi:10.18608/jla.2019.62.1.

[24] RETS I, HERODOTOU C, GILLESPIE A. Six Practical Recommendations Enabling Ethical Use of Predictive Learning Analytics in Distance Education. Journal of Learning Analytics, 2023, 10(1): 149-167. doi:10.18608/jla.2023.7743.

[25] JONES K M L, ASHER A D, GOBEN A, et al. "We're Being Tracked at All Times": Student Perspectives of Their Privacy in Relation to Learning Analytics in Higher Education. Journal of the Association for Information Science and Technology, 2020, 71(9): 1044-1059. doi:10.1002/asi.24358.

[26] MUTIMUKWE C, VIBERG O, ÖBERG L M, et al. Students' Privacy Concerns in Learning Analytics: Model Development. British

Journal of Educational Technology, 2022, 53(4): 932-951. doi:10.1111/bjet.13234.

[27] SEPP S. Towards More Transparency in Learning Analytics: Sharing Information with University Students Increases Their Awareness of Data Collection Practices. Journal of Learning Analytics, 2025, 12(3): 34-46. doi:10.18608/jla.2025.8713.

[28] DAI W, LIN J, JIN F J Y, et al. Learning Analytics for Early Identification of At-Risk Students and Feedback Intervention. Journal of Learning Analytics, 2025, 12(3): 102-125. doi:10.18608/jla.2025.8735.

[29] KIZILCEC R F, LEE H. Algorithmic Fairness in Education. In HOLMES W, PORAYSKA-POMSTA K, eds. The Ethics of Artificial Intelligence in Education: Practices, Challenges, and Debates [C]. London: Routledge, 2022: 174-202. doi:10.4324/9780429329067-10.

[30] MCCONVEY K, GUHA S. Designing for Fairness in Higher Education Early Warning Systems. In Proceedings of the Canadian Conference on Artificial Intelligence. 2024: Article 2024 GL14. doi:10.21428/594757db.c7aec3c1.

[31] FROEHLICH L, WEYDNER-VOLKMANN S. Adaptive Interventions Reducing Social Identity Threat to Increase Equity in Higher Distance Education: A Use Case and Ethical Considerations on Algorithmic Fairness. Journal of Learning Analytics, 2024, 11(2): 112-122. doi:10.18608/jla.2024.8301.

[32] DU P, LIU T, XIAN X. Automation Bias in Teachers' Evaluation of Student Writing: Effects of Algorithmic Warnings and Visual Risk Cues in AI Detection Reports. Frontiers in Psychology, 2026, 17: 1889402. doi:10.3389/fpsyg.2026.1889402.

[33] HACKING I. The Looping Effects of Human Kinds. In SPERBER D, PREMACK D, PREMACK A J, eds. Causal Cognition: A Multidisciplinary Debate [C]. Oxford: Clarendon Press, 1995: 351-383, especially 369-370.

[34] Ministry of Education. General Requirements for Personal Information Protection of Smart Education Platforms: JY/T 0643-2025. (2025-04-27) [2026-08-13]. https://hudong.moe.gov.cn/srcsite/A16/s3342/202507/W020250902563098499971.pdf.

[35] MI H Q, WU H N. Mechanisms, Dilemmas, and Approaches of AI Intelligent Agents Empowering Precision Ideological and Political Education. Ideological Education Research, 2026(3): 36-42.

[36] GONG S S, YAN X X. Research on Stratified Responsibility Attribution Mechanism under the Conceptual Construction of Educational Algorithm Subjects. Distance Education Journal, 2026(2): 63-72. doi:10.15881/j.cnki.cn33-1304/g4.2026.02.007.