

From Macro Warnings to Micro Risks: Identifying Generative AI Risks in University Ideological and Political Education

Yanhua Zhong*, Baoquan Xie, Yufeng Zou

Institute of Marxism, Ganzhou Polytechnic, Ganzhou, Jiangxi, 341000, China

*Corresponding Author

DOI: <https://doi.org/10.51583/IJLTEMAS.2026.150400048>

Received: 12 April 2026; Accepted: 17 April 2026; Published: 06 May 2026

ABSTRACT

The rapid integration of Generative Artificial Intelligence (GAI) into university ideological and political education has created new pedagogical opportunities while simultaneously introducing significant risks. While existing research has produced valuable macro level warnings concerning ideological security, discursive ethics, and teacher authority erosion, these warnings remain largely at the level of principle based alerts. They lack the granularity needed for frontline teachers to recognize and respond to specific risks in actual teaching scenarios. This study addresses this gap by shifting the analytical focus from macro warnings to micro risks. Employing a mixed methods approach that includes a survey of 500 ideological course instructors, semi structured interviews with teachers and students, and classroom observations across multiple universities in Jiangxi Province, the study systematically identifies risk manifestations across four teaching scenarios: lesson preparation, classroom instruction, interactive sessions, and assessment activities. The findings reveal concrete, observable, and intervenable micro risks, such as GAI generated case studies that subtly weaken the Party's leadership narrative, students uncritically copying AI generated answers, and automated grading systems that fail to detect ideologically problematic statements. A preliminary micro risk taxonomy is developed to make these risks visible, nameable, and actionable. The study further identifies three causal logics underlying these risks: technological bias embedded in training data, institutional absence of review mechanisms, and cognitive blind spots among teachers who mistakenly trust GAI neutrality. Theoretically, this study shifts the field from macro warnings to micro identification. Practically, the micro risk taxonomy provides a diagnostic tool that frontline teachers and administrators can use to recognize and respond to GAI risks in their daily work.

Keywords: Generative AI, risk identification, micro risks, ideological and political education, higher education

INTRODUCTION

The rapid integration of Generative Artificial Intelligence (GAI), exemplified by large language models such as ChatGPT and DeepSeek, into university education has opened new pedagogical possibilities. In the specific context of ideological and political education in Chinese universities, GAI offers promising opportunities for personalized learning, intelligent tutoring, and dynamic content generation. However, alongside these opportunities, the deployment of GAI in this value-laden educational domain has raised significant concerns regarding content security, discursive ethics, teacher authority, and data privacy. Recent global reviews have documented the rapid proliferation of GAI policies in higher education institutions, highlighting both

opportunities and governance challenges (Jin et al., 2025).

Existing research has produced valuable macro-level warnings about the risks of GAI in ideological education. Scholars have identified three dominant categories of risk: ideological security risks, such as the potential weakening of the Party's leadership narrative through algorithmically generated content (Yu, 2025); ethical and discursive risks, including the spread of historical nihilist expressions and the erosion of teacher authority as a discursive gatekeeper (Hu, 2025); and subjectivity erosion risks, where both teachers and students may fall into uncritical reliance on AI outputs (Yu, 2025; Meng & Yao, 2025). While these warnings are theoretically important and practically relevant, they share a common limitation: they remain largely at the level of principle-based alerts, lacking the granularity needed for frontline teachers to recognize and respond to specific risks in actual teaching scenarios.

This gap between macro warnings and micro practices is not merely academic. For a teacher preparing a lesson on "Reform and Opening-up," a macro warning about "ideological security risks" offers little guidance on how to detect whether a GAI-generated case study subtly weakens the leadership narrative. For a student engaging with an AI-powered Q&A tool, a macro warning about "subjectivity erosion" does not help the instructor identify when the student is uncritically copying AI-generated answers. The absence of a micro-level, scenario-specific risk identification framework leaves teachers without actionable diagnostic tools and leaves institutions without clear governance protocols.

This study addresses this critical gap by shifting the analytical focus from abstract principles to concrete, observable, and intervenable micro-risks embedded in daily pedagogical practices. Specifically, this research poses the following questions: (1) What specific risk manifestations emerge when teachers and students use GAI across different teaching scenarios in ideological and political courses? (2) How can these micro-risks be systematically identified, named, and categorized? (3) What causal logics underlie the generation of these risks?

Employing a mixed-methods approach that includes a survey of ideological course instructors, semi-structured interviews with both teachers and students, and classroom observations across multiple universities in Jiangxi Province, this study systematically identifies risk manifestations across four key teaching scenarios: lesson preparation, classroom instruction, interactive sessions, and assessment activities. Based on empirical data, the study proposes a preliminary micro-risk taxonomy for GAI in ideological education, moving the field from macro warnings toward an empirically grounded, operationally useful framework for risk identification.

The remainder of this paper is structured as follows. Section 2 reviews the existing literature on GAI risks in ideological education, highlighting the macro-to-micro gap. Section 3 presents the analytical framework and research methodology. Section 4 reports the empirical findings, including specific risk manifestations across the four teaching scenarios and a preliminary risk taxonomy. Section 5 discusses the theoretical and practical implications of shifting from macro warnings to micro identification. Section 6 concludes with limitations and directions for future research.

LITERATURE REVIEW

Three dominant macro level risk discourses

The existing literature on GAI in ideological and political education has produced three dominant categories of macro level risk warnings. A systematic review of GAI ethics in higher education has identified recurring themes of algorithmic bias, data privacy, and pedagogical role transformation (Barus et al., 2025). These categories, while theoretically important, collectively illustrate the macro level orientation that this study seeks to move

beyond.

The first category concerns ideological security risks. Scholars have warned that GAI, particularly large language models trained on diverse and often unregulated datasets, may generate content that weakens or distorts official narratives. Yu (2025) argues that algorithmic recommendations can reinforce historical nihilist tendencies by presenting decontextualized or one-sided accounts of sensitive historical events. Similarly, Hu (2025) points out that GAI-generated content may inadvertently dilute the leadership narrative of the Party when producing case studies on topics such as “Reform and Opening up” or “Common Prosperity.” These risks are not merely hypothetical. The underlying concern is that GAI systems, optimized for coherence and user engagement rather than political accuracy, may produce outputs that are factually correct but ideologically misaligned.

The second category focuses on ethical and discursive risks. Hu (2025) provides a systematic analysis of how GAI reshapes the discursive ecology of ideological courses. He identifies three core dilemmas: the dissolution of teacher authority as algorithmic recommendations dilute the teacher's role as a discursive gatekeeper, the polarization of student audiences as personalized recommendation systems reinforce cognitive echo chambers, and discursive disorder as GAI-generated content struggles to balance political correctness with factual accuracy. This analysis draws on Habermas's theory of communicative action, which emphasizes truth, sincerity, and legitimacy as preconditions for productive discourse. When GAI-generated content lacks these qualities, the entire discursive foundation of ideological education is threatened.

The third category addresses subjectivity erosion risks. Yu (2025) and Meng and Yao (2025) have documented how both teachers and students may develop uncritical dependence on GAI outputs. For teachers, the risk is becoming a “technical operator” who merely delivers AI-generated content without critical mediation. For students, the risk is falling into “cognitive outsourcing,” where they accept AI-generated answers without questioning their validity or underlying assumptions. Selwyn (2019) has similarly warned in a broader educational context that AI tools, if not carefully governed, can lead to the commodification of education and the dehumanization of teacher-student relationships. Williamson (2017) adds that learning analytics technologies risk reducing students to “data points,” undermining their integrity as moral agents.

While these three categories provide a useful starting point, they remain largely descriptive rather than analytically integrated. Few studies have examined how these risk categories interact or reinforce each other. For instance, ideological security risks may exacerbate subjectivity erosion when students uncritically accept biased AI outputs, yet this intersection remains underexplored in the existing literature.

Why macro warnings are insufficient

Despite their theoretical value, macro-level warnings share a critical limitation. They remain at the level of principle-based alerts and lack the granularity needed for frontline teachers to recognize and respond to specific risks in actual teaching scenarios.

This limitation manifests in three specific ways. First, risk identification is abstract rather than concrete. Existing studies warn about “algorithmic bias” or “ideological security risks” but rarely specify what these risks look like when a teacher is preparing a lesson, facilitating a classroom discussion, or evaluating student assignments. Second, governance recommendations are aspirational rather than operational. Scholars commonly call for “strengthening value rationality,” “enhancing ethical standards,” or “improving digital literacy” (Meng & Yao, 2025; Hu, 2025), but few provide concrete tools that teachers can directly use, such as checklists for reviewing AI-generated content or question banks for guiding student critique. Third, the research perspective is technology-centered rather than teaching-centered.

Much of the existing literature focuses on the technical logic of GAI systems themselves, treating ideological education as a generic application context rather than a distinctive value laden domain with its own pedagogical logic and political imperatives. Large scale studies on the ethical and societal impacts of GAI in higher education have similarly noted a tendency to prioritize technical affordances over pedagogical contexts (Mak et al., 2025).

The call for micro level risk research

A small but growing body of literature has begun to call for more micro level, context sensitive approaches to GAI risk research. In the field of AI ethics, Floridi (2019) has argued that digital ethics must move from principles to practices, embedding ethical requirements into technical design rather than leaving them as abstract aspirations. Bostrom and Yudkowsky (2018) have warned about the danger of “instrumental convergence,” where an AI system optimized for a narrow goal may pursue that goal in ways that violate broader human values. These insights, while developed primarily in the context of general AI safety, have direct relevance for ideological education. If GAI systems are optimized for student engagement or content coherence without explicit value alignment, they may generate outputs that are pedagogically effective but ideologically problematic.

In the specific context of Chinese ideological education, the call for micro level research is particularly urgent. The macro warnings currently available do not answer practical questions such as: When a teacher uses GAI to generate a case study on “Reform and Opening up,” what specific wording patterns should trigger a manual review? When a student uses a GAI powered Q&A tool, how can the instructor distinguish between legitimate help and uncritical copying? When an institution considers adopting a GAI system for automated grading, what privacy safeguards must be in place? Table 1 provides a structured comparison between existing macro level risk research and the micro level approach taken by this study.

Table 1. Summary of Macro Level Risk Research and the Micro Level Gap

Dimension	What Existing Research Has Done (Macro Level)	What This Study Does (Micro Level)
Risk identification	Provides principle based warnings (e.g., “ideological security risks,” “subjectivity erosion”)	Identifies concrete, observable, and intervenable risk manifestations across specific teaching scenarios
Risk categories	Offers three broad categories: ideological security, ethical and discursive, and subjectivity erosion	Develops a scenario specific micro risk taxonomy (lesson preparation, instruction, interaction, assessment)
Governance focus	Calls for strengthening value rationality, enhancing ethical standards, and improving digital literacy	Provides operational diagnostic tools including a risk checklist and identification framework
Research perspective	Technology centered, treating ideological education as a generic application context	Teaching centered, grounded in the pedagogical logic and political imperatives of ideological courses
Empirical basis	Largely theoretical or conceptual with limited classroom level data	Empirically grounded in surveys, interviews, and classroom observations across multiple universities

Research gap and contribution of this study

The literature review reveals a clear research gap. Existing studies have produced valuable macro level warnings about GAI risks in ideological education, but they have not systematically identified the specific, observable, and intervenable micro risks that emerge in actual teaching scenarios. Consequently, teachers lack actionable diagnostic tools, and institutions lack clear governance protocols.

This study addresses this gap by shifting the analytical focus from macro warnings to micro identification. The contribution is twofold. Theoretically, the study develops a preliminary micro risk taxonomy for GAI in ideological education, moving the field from abstract principle based alerts to empirically grounded, scenario specific risk categories. Practically, the study provides a diagnostic tool that frontline teachers and administrators can use to recognize, name, and respond to GAI risks in their daily work. This shift from macro to micro is a necessary step toward developing governance strategies that serve rather than alarm frontline educators.

RESEARCH METHODOLOGY

Research design

This study employs a mixed methods research design that combines quantitative and qualitative approaches. The rationale for using a mixed methods design is twofold. First, the research questions require both breadth (to understand the prevalence of GAI risks across a larger population) and depth (to capture the specific manifestations of risks in authentic classroom contexts). Second, the exploratory nature of this study, which seeks to identify previously unnamed micro risks, benefits from qualitative data that can capture unexpected phenomena while quantitative data provides a sense of scale. **Table 2** below summarizes the key components of the research design and data collection.

Table 2. Summary of Research Design and Data Collection

Research Component	Description
Research design	Mixed methods (quantitative + qualitative)
Quantitative method	Survey questionnaire
Quantitative sample	500 ideological course instructors in Jiangxi Province
Qualitative method 1	Semi structured interviews
Qualitative sample 1	20 ideological course instructors
Qualitative sample 2	25 students who have used GAI assisted learning
Qualitative method 2	Non-participatory classroom observations
Number of observations	2 lessons per participating teacher
Additional materials	GAI generated screenshots, teaching reflections, and negative cases
Data analysis tools	NVivo for qualitative (thematic analysis); SPSS for quantitative (descriptive statistics)

Quantitative component

The quantitative component consists of a survey questionnaire distributed to ideological course instructors in universities across Jiangxi Province. The target sample size is 500 respondents. The survey focuses on three core questions: whether and how frequently instructors use GAI for teaching, what specific problems they have encountered, and what types of support they most need.

The survey instrument was developed based on the literature review and pilot tested with five instructors for clarity and relevance. The final version includes closed ended questions with Likert scale responses and multiple choice items. Descriptive statistics, including frequency distributions and percentages, are used to analyze the survey data.

Qualitative component

The qualitative component includes three data collection methods: semi structured interviews, non participatory classroom observations, and collection of supplementary materials such as GAI generated screenshots and teaching reflections.

Semi structured interviews

Interviews are conducted with two groups of participants. The first group includes 20 ideological course instructors with varying years of teaching experience, professional titles, and frequencies of GAI use. The interview protocol explores their practical experiences, concerns, coping strategies, and institutional expectations regarding GAI use. The second group includes 25 students who have experienced GAI assisted instruction. The student interview focuses on their trust in AI generated content, critical awareness, and privacy perceptions. All interviews are audio recorded with participant consent and transcribed verbatim.

Classroom observations

Non participatory classroom observations are conducted in 2 lessons per participating teacher. The purpose is to document how GAI is actually used in real time, how teachers mediate AI generated content, and how students respond. Potential risk moments, such as when GAI generated content contains value biases that go unnoticed, are recorded as observation notes.

Supplementary materials

Participating teachers are invited to share GAI related teaching materials, including screenshots of AI generated content, teaching reflections on GAI use, and examples of problematic outputs. These materials serve as additional data sources for identifying risk manifestations.

Data analysis

Qualitative data from interviews, observations, and supplementary materials are analyzed using thematic analysis with the assistance of NVivo software. The analysis follows a three stage coding process: open coding to identify initial themes, axial coding to group themes into categories, and selective coding to integrate categories into a coherent framework.

Quantitative data from the survey are analyzed using SPSS. Descriptive statistics, including frequency distributions and means, are used to summarize the prevalence of different risk types and support needs. The quantitative results are used primarily to validate and contextualize the qualitative findings rather than as the main analytical driver.

Ethical considerations

This study has received ethical approval from the affiliated university. All participants are informed of the research purpose and their rights, including the right to withdraw at any time. Informed consent is obtained from all participants. For classroom observations, both the instructor and the students are informed in advance. All

data are anonymized, with personal identifiers removed before analysis.

RESULTS

This section presents the empirical findings organized around the four teaching scenarios examined in this study: lesson preparation, classroom instruction, interactive sessions, and assessment activities. The findings are drawn from survey data, interview transcripts, classroom observation notes, and supplementary materials collected from participating teachers and students.

Risk manifestations across four teaching scenarios

Lesson preparation risks

In the lesson preparation scenario, teachers reported using GAI to generate case studies, historical summaries, and discussion questions. While GAI significantly reduced preparation time, it also introduced subtle but consequential risks. Several teachers noted that GAI generated case studies on politically sensitive topics such as “Reform and Opening up” tended to emphasize economic achievements while downplaying the leadership narrative of the Party. One teacher explained, “The AI generated text was factually correct, but something felt off. It told the story of economic growth without mentioning who led the reform.” Similarly, GAI produced historical summaries occasionally contained decontextualized accounts of sensitive events, presenting them as neutral facts rather than as part of a coherent ideological framework.

Classroom instruction risks

During classroom instruction, teachers displayed GAI generated content on screens or used GAI to provide real time examples. The most frequently reported risk was the presence of historically nihilist expressions. For instance, when a teacher used GAI to generate a description of a revolutionary hero, the output described the hero’s personal struggles in detail but omitted the broader political significance of their actions. Another teacher reported that GAI generated definitions of “Common Prosperity” overemphasized efficiency and market mechanisms while underemphasizing equity and redistribution. A third risk involved implicit bias, where GAI produced examples that presented Western political systems as neutral benchmarks for comparison without acknowledging their ideological assumptions.

Interactive session risks

In interactive sessions, students used GAI powered Q&A tools to assist with in class discussions and assignment preparation. The most concerning risk was students copying AI generated answers without critical scrutiny. Interview data revealed that many students trusted GAI outputs as authoritative, rarely questioning their validity or underlying assumptions. One student admitted, “If ChatGPT gives me an answer, I assume it is correct. I do not have time to fact check everything.” This phenomenon, which can be described as cognitive outsourcing, represents a form of subjectivity erosion where students delegate critical thinking to the AI. Recent research on GAI in ideological education has similarly identified this erosion of educational object subjectivity as a core ethical concern (Yan, 2025). Additionally, classroom observations recorded instances where GAI chatbots provided vague or evasive responses to politically sensitive student questions, leaving students confused rather than informed.

Assessment activity risks

In the assessment scenario, teachers reported using GAI to assist with grading and feedback, while students used

GAI to generate essay submissions. Teachers noted that automated grading systems sometimes penalized politically appropriate but linguistically non standard expressions, suggesting algorithmic bias in how content is evaluated. Conversely, automated feedback systems occasionally failed to detect ideologically problematic statements in student submissions, creating a content blind spot. On the student side, several teachers expressed concern about academic integrity, noting that some students submitted GAI generated essays without proper disclosure or critical adaptation. As one teacher put it, “The students are not cheating in the traditional sense, but they are also not learning. The AI does the thinking for them.”

Summary of micro risk taxonomy

To provide a structured overview of the empirical findings, **Table 3** presents a micro risk taxonomy that organizes specific risk manifestations by teaching scenario and risk type. This taxonomy moves beyond macro level warnings by naming concrete, observable, and intervenable risks that frontline teachers can recognize in their daily practice.

Table 3. Micro Risk Taxonomy of GAI in Ideological and Political Education

Teaching Scenario	Specific Risk Manifestations (Examples)	Risk Type
Lesson preparation	GAI generated case studies on "Reform and Opening up" that subtly weaken the Party's leadership narrative	Content security
	GAI produced historical summaries that contain decontextualized or one sided accounts of sensitive events	Ideological misalignment
	GAI suggested discussion questions that assume value neutrality rather than explicit political framing	Discursive risk
Classroom instruction	GAI outputs containing historically nihilist expressions when describing revolutionary heroes	Ideological security
	GAI generated definitions of "Common Prosperity" that overemphasize efficiency while underemphasizing equity	Value distortion
	GAI produced examples that present Western political systems as neutral benchmarks for comparison	Implicit bias
Interactive sessions	Students copying GAI generated answers without critical scrutiny or source attribution	Subjectivity erosion
	Students accepting GAI outputs as authoritative without questioning underlying assumptions	Cognitive outsourcing
	GAI chatbots providing vague or evasive responses to politically sensitive student questions	Discursive disorder
Assessment activities	GAI assisted grading that penalizes politically appropriate but linguistically non standard expressions	Algorithmic bias
	Automated feedback systems that fail to detect ideologically problematic student statements	Content blind spot
	Student submission of GAI generated essays without proper disclosure or critical adaptation	Academic integrity

Causal logics underlying micro risks

Beyond identifying specific risk manifestations, this study also examined why these risks occur. The data suggest three interconnected causal logics.

Technological bias

The first causal logic concerns the training data and optimization objectives of GAI systems. Many GAI models are trained on diverse but predominantly Western datasets, which embed certain political and cultural assumptions. When these models are applied to the Chinese ideological education context, they may produce outputs that are statistically coherent but ideologically misaligned. Additionally, GAI systems are optimized for user engagement and content fluency rather than political accuracy, creating a built-in tension with the requirements of ideological education.

Institutional absence

The second causal logic involves the lack of institutional policies and review mechanisms. Most participating teachers reported that their universities had no clear guidelines on acceptable GAI use in teaching. Without institutional standards, individual teachers are left to navigate GAI risks on their own, leading to inconsistent practices. One teacher explained, "I do not know what is allowed and what is not. I just try to be careful, but I have no official guidance."

Cognitive blind spots

The third causal logic relates to cognitive blind spots among teachers. Several interview participants expressed misplaced trust in GAI neutrality, assuming that AI-generated content is inherently objective or value-free. This assumption overlooks the embedded values and biases in GAI systems. As one teacher admitted, "I never thought that AI could have a political bias. I assumed it was just pulling facts from the internet" (Interview, Teacher #12, 2025). This cognitive blind spot leaves teachers vulnerable to unrecognized ideological misalignments in GAI outputs.

DISCUSSION

This section discusses the theoretical and practical implications of shifting the analytical focus from macro warnings to micro risks in the study of GAI in ideological and political education. It also addresses the study's limitations and proposes directions for future research.

The macro to micro shift as a necessary paradigm turn

The findings of this study demonstrate that macro-level warnings, while theoretically important, are insufficient for frontline teachers who need to recognize and respond to specific GAI risks in real time. By identifying concrete risk manifestations across four teaching scenarios, this study provides empirical evidence that micro risks are not merely smaller versions of macro risks. Rather, they represent a distinct analytical level that requires its own vocabulary, taxonomy, and diagnostic tools.

This macro-to-micro shift parallels similar developments in other fields. In public health, for example, macro-level warnings about "health risks" are useful but must be supplemented by micro-level screening tools that help clinicians identify specific symptoms. In cybersecurity, general warnings about "data breaches" are less useful than specific checklists that help users recognize phishing attempts. Similarly, in GAI risk research, moving from principle-based alerts to scenario-specific risk identification is a necessary step toward developing operable governance strategies.

The theoretical contribution of this shift is twofold. First, it challenges the assumption that macro-level warnings naturally translate into micro-level practices. The gap documented in this study suggests that translation is neither

automatic nor straightforward. Second, it provides an empirically grounded micro risk taxonomy that can serve as a foundation for future theory building. Rather than starting from abstract principles, future research can build on the specific risk types identified in this study to develop more nuanced models of GAI risk generation and propagation.

Redefining teacher professionalism in the GAI era

The findings also have implications for how teacher professionalism is understood in the context of GAI. Much of the existing literature has focused on the threat that GAI poses to teacher authority, framing the relationship between teachers and AI as inherently competitive. This study suggests a different framing. The teachers in this study were not passive victims of technological disruption. They were actively developing strategies to review, critique, and guide GAI use, even in the absence of institutional support.

This finding aligns with the concept of teacher agency, which emphasizes the active role that teachers play in shaping technology use rather than being shaped by it. In the GAI era, teacher professionalism may need to be redefined to include not only content knowledge and pedagogical skill but also what might be called algorithmic discernment: the ability to identify value misalignments in AI generated content, to scaffold student critique of AI outputs, and to make contextual judgments about when and how to use GAI in politically sensitive teaching environments.

This redefinition has practical implications for teacher training and professional development. Pre service and in service training programs should move beyond generic digital literacy modules to include scenario based training on GAI risk identification. Teachers need opportunities to practice reviewing GAI generated content, to discuss edge cases with peers, and to develop shared norms for acceptable GAI use.

From individual coping to institutional governance

A third implication concerns the level at which GAI risks are best governed. The findings reveal that most teachers are currently coping with GAI risks on an individual basis, with little institutional guidance or support. This individual coping strategy is unsustainable and inequitable. Teachers with higher digital literacy or more institutional support are better equipped to manage risks, while others are left vulnerable.

The data suggest a need for institutional governance mechanisms that operate at three levels. At the teacher level, institutions should provide practical tools such as risk checklists, review protocols, and teaching scripts that help teachers identify and respond to common GAI risks. At the course level, institutions should support curriculum integration efforts that embed AI literacy and critical AI use into existing course structures rather than treating them as add ons. At the institutional level, universities should develop clear policies on acceptable GAI use, establish review mechanisms for high risk applications such as automated grading, and create reporting channels for problematic GAI outputs. This three level governance framework aligns with recent institutional intervention models proposed in the literature (Jayasinghe et al., 2026).

Notably, these governance mechanisms need not be burdensome or expensive. Many of the tools that teachers requested in this study, such as a one page risk checklist or a set of example teaching scripts, are low cost interventions that could be developed and disseminated with modest resources. The barrier is not primarily financial but organizational, requiring institutions to prioritize GAI governance as part of their teaching quality assurance systems.

While this study focuses on risk identification, it is important to note that identification is only the first step toward mitigation. Future research should build on the taxonomy developed here to design and test specific

intervention strategies. Potential mitigation approaches include teacher training modules on AI output review, student workshops on critical AI literacy, and institutional policy frameworks for GAI governance. The taxonomy provides a diagnostic foundation for these interventions by specifying what risks need to be addressed.

Limitations

This study has several limitations that should be acknowledged. First, the empirical data were collected exclusively from universities in Jiangxi Province. While this regional focus allowed for in depth data collection, it limits the generalizability of the findings to other provinces or countries with different political, cultural, or technological contexts. Second, the study relied primarily on self reported data from interviews and surveys, which may be subject to social desirability bias or recall limitations. Classroom observations helped mitigate this limitation, but observational data were limited to one or two lessons per teacher. Third, the rapid pace of GAI development means that specific risk manifestations may change as new models and applications emerge. The taxonomy proposed in this study should be treated as a preliminary framework that requires regular updating rather than a fixed classification. Fourth, this study focused on risk identification rather than risk mitigation or intervention testing. While the taxonomy provides a diagnostic foundation, future research is needed to develop and test specific interventions.

Future research directions

The findings of this study open several avenues for future research. First, the micro risk taxonomy should be validated and refined through larger scale studies across multiple provinces and institutional types. Quantitative surveys could assess the prevalence of different risk types and identify teacher or institutional characteristics associated with higher risk exposure. Second, intervention research is needed to test the effectiveness of specific governance tools, such as risk checklists or training programs. Randomized controlled trials or quasi experimental designs could compare outcomes between teachers who receive specific interventions and those who do not. Third, comparative research across different countries or political systems would help clarify which aspects of GAI risk are universal and which are context specific. Such research would contribute to a more nuanced understanding of the relationship between GAI, ideology, and education. Fourth, longitudinal research could track how GAI risks evolve over time as technologies change and as teachers and students develop new coping strategies. Such research would help anticipate emerging risks before they become widespread.

CONCLUSION

This study addressed the gap between macro level warnings and micro level practices in research on Generative Artificial Intelligence (GAI) in university ideological and political education. Drawing on survey data, interviews, and classroom observations across multiple universities in Jiangxi Province, the study identified specific, observable, and intervenable risk manifestations across four teaching scenarios: lesson preparation, classroom instruction, interactive sessions, and assessment activities. A preliminary micro risk taxonomy was developed to make these risks visible, nameable, and actionable. Theoretically, the study shifts the analytical focus from macro warnings to micro identification. Practically, the taxonomy provides a diagnostic tool for frontline teachers and administrators.

The study has several limitations, including its regional focus and its reliance on self-reported data, and its emphasis on risk identification rather than intervention testing. Additionally, given the exploratory phase of this research and the novelty of the micro-risk taxonomy, the quantitative analysis was primarily descriptive. This approach is appropriate for establishing foundational patterns and frequencies. Future research, building on this taxonomy, should employ inferential statistical techniques (e.g., regression analysis) to examine the predictive

relationships between teacher characteristics, institutional contexts, and the prevalence of specific micro-risks. The integration of GAI into ideological education is unlikely to reverse. The question is not whether to use these technologies, but how to govern their risks intentionally. This study has taken a first step toward answering that question.

REFERENCES

1. Barus, O., Hidayanto, A. N., & Eitiveni, I. (2025). Mapping generative AI's ethical issues in higher education: A FELT-guided systematic review. *Polyglot: Jurnal Ilmiah*, 21(2). <https://doi.org/10.19166/pji.v21i2.10020>
2. Bostrom, N., & Yudkowsky, E. (2018). The ethics of artificial intelligence. In *Artificial intelligence safety and security* (pp. 57-69). Chapman and Hall/CRC.
3. Dai, J. P., & Qin, Y. Y. (2023). The ideological risks of generative artificial intelligence such as ChatGPT and its response. *Journal of Chongqing University (Social Science Edition)*, 29(5), 101-110.
4. Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being ethical. *Philosophy & Technology*, 32(2), 185-193. <https://doi.org/10.1007/s13347-019-00354-x>
5. Habermas, J. (1984). *The theory of communicative action* (T. McCarthy, Trans.). Beacon Press. (Original work published 1981)
6. Hu, G. (2025). Discursive ethical risks and governance paths of ideological and political courses in universities in the era of generative artificial intelligence. *Heilongjiang Researches on Higher Education*, 43(9). (forthcoming)
7. Jin, Y., Yan, L., Echeverria, V., Gašević, D., & Martinez-Maldonado, R. (2025). Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. *Computers and Education: Artificial Intelligence*, 8, 100348. <https://doi.org/10.1016/j.caeai.2024.100348>
8. Mak, J., Nakatumba-Nabende, J., Clear, T., Clear, A., Albluwi, I., Andrei, O., Angeli, L., MacNeil, S., Oyelere, S. S., Rattigan, M. H., Sheard, J., & Zhu, T. (2025). Navigating the ethical and societal impacts of generative AI in higher computing education (arXiv:2511.15768v1). arXiv. <https://doi.org/10.48550/arXiv.2511.15768>
9. Meng, Q. P., & Yao, H. X. (2025). The internal mechanism, risks and countermeasures of AI-driven teaching reform in university ideological and political courses. *Modern Distance Education Research*, 37(3). (forthcoming)
10. Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.
11. Jayasinghe, S., Gamage, K. A., Yang, D., Cheng, C., Disanayake, C., & Apeji, U. D. (2026). Six Institutional Intervention Areas to Support Ethical and Effective Student Use of Generative AI in Higher Education: A Narrative Review. *Education Sciences*, 16(1), 137.. <https://doi.org/10.3390/educsci16010137>
12. Wang, S. J., & Zhang, Y. (2024). The basic logic and contradiction adaptation of generative AI intervening in ideological and political education: From ChatGPT to GPT-4o. *Ideological Education Research*, 2024(12), 52-58.
13. Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. SAGE Publications.
14. Yan, R. F. (2025). Ethical risks and resolution paths of generative AI empowering ideological and political education: An investigation based on the perspective of educational object subjectivity. *Marxist Studies Network*. http://marxism.cass.cn/zzjy/202510/t20251029_5921859.shtml
15. Yu, Y. (2025). Risk prevention and practical exploration of AI-empowered ideological and political course teaching in universities. *Journal of Langfang Normal University (Social Sciences Edition)*, 41(2).
16. Yue, Q., & Chen, M. Z. (2025). Coupled mechanisms, risk challenges, and ecological reconstruction of

GenAI enabling ideological and political education. *Journal of Zhengzhou University of Light Industry (Social Science Edition)*, 26(5), 59-67. <https://doi.org/10.12186/2025.05.008>

17. Ludi, Z. H. A. O., & Shengli, Q. I. (2025). Ethical Risks and Prevention of Generative Artificial Intelligence Empowering Ideological and Political Education for College Students. *The journal of xinyang normal university (philosophy and social science edition)*, 45(6), 16-24.
18. Zhong, H. (2025). The Opportunities, Challenges and Countermeasures Brought by the Development of Chat Generative. In *Proceedings of the 2025 9th International Seminar on Education, Management and Social Sciences (ISEMSS 2025)* (p. 281). Springer Nature. https://doi.org/10.2991/978-2-38476-462-4_32