INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN PROGRESSIVE EDUCATION & DEVELOPMENT

OPEN ACCESS JOURNAL

Vol 15, Issue 2, (2026) E-ISSN: 2226-6348

The Teacher as Gatekeeper: Identifying and Governing Generative AI Risks in Chinese University Ideological and Political Courses

Yanhua Zhong, Juan Huang, Haoran Hu

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

Email: zhongyanhua@graduate.utm.my, huangjuan@gzpt.edu.cn, huhaoran@gzpt.edu.cn

DOI Link: http://dx.doi.org/10.6007/IJARPED/v15-i2/28108

Published Online: 30 April 2026

Abstract

The rapid adoption of Generative Artificial Intelligence (GAI) in Chinese university ideological and political courses has placed frontline teachers in a challenging position. While existing research has focused on macro-level risk warnings, little is known about how teachers can actively govern these risks in their daily practice. This study addresses this gap by conceptualizing the teacher as a "gatekeeper" who filters, critiques, and guides GAI use in value-laden educational contexts. Drawing on semi-structured interviews with 20 ideological course instructors and classroom observations across five universities in Jiangxi Province, the study identifies four gatekeeping strategies that teachers spontaneously develop: AI output review, critical questioning in class, boundary setting, and peer consultation. The findings also reveal three governance dilemmas faced by teachers: the lack of standards, time constraints, and institutional policy gaps. Based on these findings, the paper proposes a three-level governance framework operating at the teacher, course, and institution levels. Unlike existing studies that emphasize technological risks or teacher replacement, this framework empowers teachers to become effective gatekeepers without being overwhelmed by technological demands. The study contributes a teacher-centered perspective to GAI governance in politically sensitive educational settings and offers practical tools, including a risk review checklist and teaching scripts.

Keywords: Generative AI, Teacher Gatekeeper, Risk Governance, Ideological Education, Qualitative Study

Introduction

The rapid integration of Generative Artificial Intelligence (GAI) into Chinese university ideological and political courses has created both opportunities and challenges for frontline teachers. Since the public release of ChatGPT in late 2022, GAI tools have been increasingly adopted in higher education settings worldwide. In China, domestic large language models such as DeepSeek and ERNIE have further accelerated this trend, making GAI accessible to teachers and students across the country (Jin et al., 2025). While these tools can assist with lesson preparation, content generation, and student interaction, they also introduce significant risks that are particularly acute in politically sensitive educational contexts.

The importance of investigating this topic cannot be overstated. Ideological and political education serves as the cornerstone for cultivating socialist core values and maintaining ideological security in Chinese universities. Any disruption or distortion in this domain may have far-reaching consequences for students' value formation and the overall effectiveness of the educational system. As GAI becomes increasingly embedded in daily teaching practices, understanding how to govern its associated risks is not merely an academic exercise but an urgent practical necessity. Without systematic risk governance, the very integrity of ideological education could be compromised, potentially undermining the foundational mission of higher education in China.

The risks associated with GAI in ideological education are well documented in the emerging literature. These include content that subtly weakens official narratives, historically nihilist expressions that distort the presentation of revolutionary events, and students uncritically accepting AI-generated answers as authoritative (Hu, 2025; Meng & Yao, 2025; Yu, 2025). Scholars have also warned about algorithmic bias embedded in training data, the erosion of teacher authority as a discursive gatekeeper, and the potential for cognitive outsourcing where students delegate critical thinking to AI systems (Crawford, 2021; Selwyn, 2019; Williamson, 2017).

Existing research has produced valuable macro-level warnings about these risks. However, a critical gap remains. Most studies focus on identifying risks rather than governing them. Scholars have called for strengthening value rationality, enhancing ethical standards, and improving digital literacy (Meng & Yao, 2025; Hu, 2025), but few have provided concrete, actionable strategies that teachers can use in their daily practice. The question of how teachers can actively govern GAI risks remains largely unanswered. As Floridi (2019) has argued in the broader context of digital ethics, moving from principles to practices is the central challenge of our time. This insight applies directly to GAI governance in ideological education.

The specific need addressed by this study is threefold. First, frontline teachers lack practical guidance on what constitutes acceptable GAI use and how to identify problematic AI-generated content in real time. Second, university administrators have no clear framework for developing institutional policies or support mechanisms for GAI governance. Third, the existing literature offers abstract risk warnings without translating them into actionable strategies. This study directly responds to these needs by shifting the focus from risk identification to risk governance and by providing empirically grounded, practical tools for teachers and institutions.

This gap is not merely academic. Frontline teachers are expected to use GAI tools while simultaneously safeguarding the ideological integrity of their courses. Yet they receive little institutional guidance on what constitutes acceptable GAI use, how to review AI-generated content for political correctness, or how to respond when students misuse these tools. The absence of clear governance frameworks places an unfair burden on individual teachers. As one teacher in this study explained, "I do not know what is allowed and what is not. I just try to be careful, but I have no official guidance." This statement captures the predicament of many teachers navigating GAI risks without institutional support.

The relevance of this study extends to multiple beneficiaries. For teachers, the findings offer concrete gatekeeping strategies that can be immediately applied in their classrooms. For university administrators, the proposed three-level governance framework provides a roadmap for developing institutional policies and support systems. For curriculum developers, the study offers guidance on embedding AI literacy into existing course structures. For policymakers, the findings inform the design of ethical guidelines and quality assurance mechanisms for GAI use in higher education. Ultimately, the primary beneficiary is the student, who deserves an educational environment where technology enhances rather than undermines value formation and critical thinking.

This study addresses this gap by shifting the focus from risk identification to risk governance. Specifically, it conceptualizes the teacher as a "gatekeeper" who actively filters, critiques, and guides GAI use in the classroom. The gatekeeper concept, which originates from media studies, provides a useful analytical lens for understanding how teachers control the flow of information in technology-mediated learning environments. In the GAI era, teachers must develop new gatekeeping skills that go beyond traditional content selection to include the critical evaluation of algorithmically generated materials.

The research poses three questions: (1) What gatekeeping strategies do teachers spontaneously develop to govern GAI risks? (2) What governance dilemmas do they face in their daily practice? (3) How can these strategies be systematized into a practical governance framework that institutions can implement?

Drawing on semi-structured interviews with 20 ideological course instructors and classroom observations across five universities in Jiangxi Province, this study identifies four gatekeeping strategies and three governance dilemmas. Based on these findings, it proposes a three-level governance framework operating at the teacher, course, and institution levels. Unlike existing studies that emphasize teacher vulnerability or replacement, this framework centers teacher agency and provides concrete, operational tools.

The remainder of this paper is structured as follows. Section 2 reviews the existing literature on GAI risks in ideological education and the underdeveloped research on teacher roles in GAI governance. Section 3 describes the qualitative research methodology, including participant recruitment, data collection procedures, and analytical approach. Section 4 presents the empirical findings, detailing the four gatekeeping strategies and three governance dilemmas. Section 5 discusses the theoretical and practical implications of these findings and proposes a three-level governance framework. Section 6 concludes with limitations and directions for future research.

Literature Review

GAI Risks in Ideological Education: Three Categories

The literature on GAI in ideological and political education has grown rapidly since 2023. While early studies focused on the potential benefits of AI for personalized learning and teaching efficiency, a growing body of research has turned attention to risks. Three dominant categories of risk have emerged from this literature.

The first category concerns ideological security risks. Scholars have warned that GAI systems, particularly large language models trained on diverse and often unregulated datasets, may generate content that weakens or distorts official narratives. Meng and Yao (2025) argue 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. Dai and Qin (2023) have further noted that the training data of mainstream GAI models predominantly reflect Western political and cultural assumptions, creating an inherent tension when these models are applied to Chinese educational contexts.

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 (1984) 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. Yan (2025) extends this analysis by examining how GAI challenges the subjectivity of educational objects, arguing that students may lose their capacity for critical engagement when they interact primarily with algorithmically generated content.

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. Wang and Zhang (2024) further note that as GAI evolves from ChatGPT to GPT-4o, the boundary between human-generated and AI-generated content becomes increasingly blurred, intensifying concerns about subjectivity erosion.

The Teacher Role in GAI Governance: An Underdeveloped Area

While the risks of GAI in ideological education are increasingly well documented, the literature on teacher roles in GAI governance remains underdeveloped. Most existing studies focus on what teachers should not do rather than what they can do. Scholars warn against over-reliance on GAI, loss of authority, and uncritical acceptance of AI outputs. However, few studies explore how teachers can actively govern these risks through their daily pedagogical practices.

This gap reflects a broader pattern in educational technology research. As Selwyn (2019) has observed, discussions of AI in education often center on technological affordances and risks, with the teacher positioned as either a beneficiary or a victim of technological change. The active role that teachers play in shaping technology use is frequently overlooked. This is particularly problematic in the context of ideological education, where teacher mediation is essential for maintaining the political and pedagogical integrity of the curriculum.

A small but growing body of literature has begun to address this gap. Floridi (2019) argues that digital ethics must move from principles to practices, embedding ethical requirements into daily operations rather than leaving them as abstract aspirations. Jayasinghe et al. (2026) propose six institutional intervention areas to support ethical and effective student use of GAI in higher education, including policy development, training programs, and technical infrastructure. 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.

However, these insights remain at a general level and have not been operationalized for the specific context of Chinese ideological education. The question of how teachers can serve as effective gatekeepers in this context requires empirical investigation grounded in the lived experiences of frontline teachers.

The Gatekeeper Concept as an Analytical Lens

The concept of a "gatekeeper" originates from media studies, where it refers to individuals who control the flow of information by determining what content is selected, emphasized, or excluded (Lewin, 1947). In educational contexts, teachers have always served as gatekeepers, selecting what content to present, how to frame discussions, what questions to ask, and which student contributions to highlight. This gatekeeping role is particularly pronounced in ideological education, where the selection and framing of content carry political significance. GAI introduces new challenges to this traditional gatekeeping role. When content is generated by algorithms rather than selected from textbooks or curated by teachers, teachers must develop new gatekeeping skills. These include reviewing AI outputs for ideological alignment, guiding students to critique AI-generated content rather than accept it passively, setting boundaries for appropriate GAI use, and making contextual judgments about when and how to integrate GAI into instruction.

This study adopts the gatekeeper concept as an analytical lens to understand how teachers govern GAI risks. Unlike existing studies that emphasize teacher vulnerability or replacement, this study focuses on teacher agency and the practical strategies teachers develop to maintain their gatekeeping role in the GAI era. The gatekeeper concept provides a framework for identifying, categorizing, and systematizing these strategies, moving the field from abstract warnings about risk to concrete guidance for governance.

Research Methodology

Research Design

This study employs a qualitative research design focused on capturing the lived experiences and practical strategies of frontline teachers. A qualitative approach is appropriate for this study for three reasons. First, the research questions seek to understand how teachers govern GAI risks in real-world contexts, a phenomenon that requires depth of understanding rather than breadth of measurement. Second, the exploratory nature of this study, which aims to identify previously undocumented gatekeeping strategies, benefits from the flexibility and openness of qualitative methods. Third, qualitative approaches are particularly well suited for capturing teacher voice and agency, which are central to this study's conceptual framework.

Participants

Participants were recruited from five universities in Jiangxi Province, including one comprehensive university, two normal universities, and two vocational colleges. This diversity of institutional types was intentionally designed to capture variation in teacher experiences and institutional contexts.

A total of 20 ideological course instructors participated in semi-structured interviews. Participants were recruited through purposive sampling, with the goal of capturing variation in teaching experience, professional title, and frequency of GAI use. The sample included 12 female and 8 male teachers. Teaching experience ranged from 3 to 25 years (mean = 11.4 years). Professional titles included 5 lecturers, 9 associate professors, and 6 professors. All participants had used GAI tools such as ChatGPT or DeepSeek for teaching purposes at least once, with 14 reporting weekly use and 6 reporting monthly use.

Data Collection

Data were collected through two primary methods.

Semi-Structured Interviews

Semi-structured interviews lasting 45 to 75 minutes were conducted with each of the 20 participants. The interview protocol was developed based on the literature review and pilot tested with two teachers who were not included in the final sample. The protocol covered four main areas: (1) participants' experiences with GAI in teaching, including specific examples of use; (2) strategies for managing GAI-related risks; (3) challenges and dilemmas encountered; and (4) suggestions for institutional support and policy development.

All interviews were conducted in Mandarin Chinese by the first author, audio recorded with participant consent, and transcribed verbatim. Transcripts were then translated into English by a professional translator, with back-translation used to verify accuracy.

Classroom Observations

Non-participatory classroom observations were conducted for a subset of 8 teachers who represented the range of teaching experience, institutional type, and GAI use frequency in the sample. Each teacher was observed for two lessons, resulting in 16 observation sessions. Each lesson lasted 45 to 50 minutes.

During observations, the observer sat at the back of the classroom and took field notes on how GAI was used (if at all), how teachers mediated AI-generated content, how students responded, and any observable risk moments. Observation notes were expanded within 24 hours of each session. In cases where teachers did not use GAI during the observed lessons, the observer discussed with the teacher after class to understand their reasons and typical practices.

Data Analysis

Interview transcripts and observation notes were analyzed using thematic analysis, following the six-phase approach outlined by Braun and Clarke (2006). The analysis proceeded through three main stages.

First, open coding was used to identify initial themes across the dataset. The first author coded all transcripts and observation notes, generating 147 initial codes. A second coder independently coded 20% of the data to assess inter-coder reliability, achieving 84% agreement. Disagreements were resolved through discussion.

Second, axial coding was used to group related codes into broader categories. Through iterative comparison, four categories of gatekeeping strategies and three categories of governance dilemmas emerged. These categories were refined through discussion among the research team.

Third, selective coding was used to integrate these categories into a coherent framework. The three-level governance framework presented in this paper represents the outcome of this integration process.

NVivo software was used to manage the coding process. To ensure trustworthiness, findings were member-checked with five participants, who confirmed that the identified strategies and dilemmas accurately reflected their experiences. Peer debriefing was also conducted with two researchers not involved in the study.

Ethical Considerations

Ethical approval for this study was obtained from the research ethics committee of the affiliated university. All participants provided written informed consent prior to participation. Participants were informed of their right to withdraw at any time without consequence. Anonymity was guaranteed, with pseudonyms used in all reporting and no identifying information included in any publication. For classroom observations, both instructors and students were informed in advance, and consent was obtained from the instructor. No student-level data were collected.

Findings

This section presents the empirical findings organized around two main themes: the gatekeeping strategies that teachers spontaneously develop to govern GAI risks, and the governance dilemmas they face in their daily practice.

Gatekeeping Strategies

Teachers in this study developed four gatekeeping strategies to govern GAI risks. These strategies emerged from their daily practice rather than from formal training or institutional guidance. Each strategy is described below, with representative quotes from participants.

Strategy 1: AI Output Review

The most common strategy reported by participants was a systematic review of AI-generated content before classroom use. Teachers consistently stated that they never used GAI outputs directly. Instead, they developed informal but rigorous review protocols. One teacher with 15 years of experience explained:

"I always run the AI output through three checks. First, factual accuracy: Are the dates, names, and events correct? Second, political alignment: Does the framing align with official narratives? Third, language appropriateness: Is the wording suitable for a classroom setting? If any check fails, I revise or discard it. This takes time, but I cannot afford to make a mistake." (Teacher #07, associate professor)

Teachers varied in the thoroughness of their review. Some used mental checklists similar to the one described above, while others relied more heavily on intuition developed through years of teaching experience. However, all participants agreed that some form of review was necessary. As another teacher noted:

"The AI is a helpful assistant, but it makes mistakes. Last week, I asked it to generate a summary of the Cultural Revolution for background context. The output was completely one-sided, focusing only on economic disruption without mentioning any of the official framing about learning from mistakes. If I had used it directly, there would have been serious problems. I caught it only because I know the content well." (Teacher #12, professor)

Strategy 2: Critical Questioning in Class

The second strategy involved using GAI outputs as teaching opportunities rather than simply discarding problematic content. Several teachers described deliberately showing students an AI-generated response that contained subtle bias or incomplete framing and then guiding them to identify and critique the problem.

One teacher provided a detailed example:

"I asked ChatGPT to explain the difference between Chinese democracy and Western democracy. The answer was technically correct in terms of factual information, but I noticed it framed Western democracy as the neutral benchmark against which Chinese democracy was compared. This is a subtle bias. So I showed the response to my students and asked them, 'What assumption is the AI making here? How can you tell?' This sparked a fantastic discussion about how AI systems embed values, how training data shapes outputs, and what it means to be a critical consumer of AI-generated content." (Teacher #03, lecturer)

This strategy transforms risk from a threat into a teachable moment. Rather than hiding GAI risks or avoiding GAI use altogether, teachers who employ this strategy use GAI outputs to develop students' critical AI literacy. As another teacher explained:

"My goal is not to prevent students from using AI. That is impossible and probably not even desirable. My goal is to teach them how to use AI critically. So when I see a problematic AI output, I do not just delete it. I bring it to class and ask, 'What is wrong with this? How would you improve it?' This way, students learn to question AI rather than trust it blindly." (Teacher #14, associate professor)

Strategy 3: Boundary Setting

The third strategy involved establishing clear boundaries for appropriate GAI use. Teachers distinguished between acceptable and unacceptable uses of GAI in their courses. While there was some variation across teachers, a general consensus emerged on several points.

Acceptable uses included: generating discussion questions to stimulate critical thinking, summarizing complex texts to make them more accessible, providing multiple examples of a concept, and assisting with language polishing for non-native speakers. Unacceptable uses included: writing complete student essays, making value judgments on politically sensitive topics, answering politically sensitive questions without teacher mediation, and generating content that substitutes for teacher expertise.

One teacher explained his boundary system in detail:
"I tell my students on the first day: You can use AI for research assistance. You can ask it to explain concepts you find difficult. You can use it to generate ideas for your own writing. But you cannot copy-paste AI answers into your assignments. If you use AI, you must cite it just like any other source. And you must include a reflection paragraph explaining how you used the AI, what you changed, and why. This makes the use of AI transparent and forces students to engage critically rather than just outsourcing their thinking." (Teacher #19, professor)

Strategy 4: Peer Consultation

The fourth strategy involved informal consultation with colleagues. Teachers reported regularly discussing GAI-related challenges with peers, sharing examples of problematic outputs, and collectively developing responses to emerging issues. This strategy emerged from necessity rather than design, as formal institutional support was largely absent.

"There is no official training or guidance at my university," one teacher noted. "So we have to figure this out together. My colleagues and I have a WeChat group where we share problematic AI outputs, ask for advice, and discuss what works. It is not ideal, but it is better than working alone." (Teacher #05, lecturer)

Peer consultation served multiple functions. It provided emotional support, helping teachers feel less isolated in their struggles with GAI governance. It distributed the workload of staying updated on rapidly evolving GAI technologies, as teachers could share new developments and responses. And it helped develop shared norms across the institution, as teachers negotiated what constituted acceptable practice in their specific context.

Table 1
Four Gatekeeping Strategies and Their Characteristics

Strategy	Description	Teacher Quote
AI output review	Systematic review of GAI-generated content before classroom use	"I always run the AI output through three checks: factual accuracy, political alignment, and language appropriateness."
Critical questioning	Using GAI outputs as teachable moments to develop student AI literacy	"I showed this to my students and asked, 'What assumption is the AI making here?'"
Boundary setting	Establishing clear rules for acceptable and unacceptable GAI use	"Students can use AI for research assistance, but they cannot copy-paste answers."
Peer consultation	Informal collaboration with colleagues to share strategies and concerns	"There is no official training or guidance. We have to figure this out together."

Governance Dilemmas

Despite developing these strategies, teachers faced three persistent governance dilemmas. These dilemmas were not failures of individual teachers but rather symptoms of systemic gaps in institutional support and policy.

Dilemma 1: Lack of Standards

The most frequently cited dilemma was the absence of clear institutional standards for GAI use. Teachers did not know what constituted acceptable GAI use, what content required mandatory human review, what consequences applied for misuse, or what documentation was required. This ambiguity placed an unfair burden on individual teachers, who had to make high-stakes judgments without guidance.

One teacher expressed frustration:
"I spend hours reviewing AI outputs because I am afraid of making a mistake. But my colleague down the hall uses AI directly in his classroom without any review. Who is right? There is no policy to tell us. We are both just guessing. And if something goes wrong, I am not sure either of us would be protected." (Teacher #11, associate professor)

This lack of standards created inconsistency across teachers and courses. Students in one class might receive strict guidance on GAI use, while students in another class received none. This inconsistency was not only unfair to students but also undermined the overall integrity of the educational program.

Dilemma 2: Time Constraints

The second dilemma involved time. Systematic review of AI-generated content was time-consuming, yet teachers had no reduction in their other responsibilities to accommodate this new task. Teachers reported spending 30 to 60 minutes reviewing and revising content that GAI generated in seconds. While they recognized the necessity of review, they worried about the long-term sustainability of this practice.

"I already have too much to do," one teacher explained. "I teach four courses, serve on two committees, advise 30 students, and am expected to publish research. Adding AI review to my workload is exhausting. But skipping review is too risky. I feel trapped between two bad options." (Teacher #02, associate professor)

This dilemma was particularly acute for early-career teachers, who faced additional pressures related to promotion and tenure. As one lecturer noted:

"Junior faculty like me are under enormous pressure to publish and teach well. Spending an extra hour per lesson reviewing AI outputs is not sustainable. But if I skip review and something goes wrong, my career could be damaged. I do not see a good solution." (Teacher #08, lecturer)

Dilemma 3: Institutional Policy Gaps

The third dilemma concerned the absence of institutional policies and support mechanisms. Most participating universities had no guidelines on GAI use in teaching, no training programs, no review mechanisms, and no reporting channels for problematic outputs. Teachers were left to navigate GAI risks entirely on their own.

One teacher summarized the situation:

"The university encourages us to use new technology. They see AI as an opportunity for innovation. But they provide no guidance on how to use it safely. No training. No policies. No support. We are experimenting on our students with no safety net. It feels irresponsible, but what are we supposed to do? Refuse to use AI and be seen as behind the times?" (Teacher #16, professor)

This policy gap extended beyond teaching to include research and assessment. Teachers reported uncertainty about whether students were using AI to complete assignments, how to detect AI-generated submissions, and what to do when they suspected AI misuse. Without institutional policies, each teacher had to develop their own approach, leading to inconsistency and confusion.

Table 2
Three Governance Dilemmas Faced by Teachers

Dilemma	Description	Teacher Quote
Lack of standards	No clear institutional guidelines on acceptable GAI use	"I spend hours reviewing AI outputs because I am afraid of making a mistake. But my colleague down the hall uses AI directly without review. Who is right?"
Time constraints	A systematic review is time-consuming and unsustainable	"Adding AI review to my workload is exhausting. But skipping the review is too risky. I feel trapped."
Institutional policy gaps	Absence of review mechanisms, training programs, and reporting channels	"The university encourages us to use new technology, but they provide no guidance on how to use it safely."

Discussion

This section discusses the theoretical and practical implications of the findings, proposes a three-level governance framework, addresses limitations, and suggests directions for future research.

Four Gatekeeping Strategies as Teacher Agency

The four strategies identified in this study demonstrate that teachers are not passive victims of technological disruption. Rather, they actively develop practices to govern GAI risks, drawing on their pedagogical expertise, content knowledge, and professional judgment. This finding challenges the dominant narrative in existing literature, which often frames teachers as vulnerable to replacement or deskilling (Selwyn, 2019; Williamson, 2017).

The strategies vary in their focus and function. AI output review and boundary setting are preventive strategies that reduce risk exposure before it occurs. These strategies operate at the level of content selection and rule setting, establishing parameters for acceptable GAI use. Critical questioning is a responsive strategy that transforms risk into a learning opportunity. Rather than simply avoiding risk, teachers who employ this strategy use GAI outputs to develop students' critical AI literacy. Peer consultation is a collaborative strategy that distributes governance work across the teacher community, leveraging collective expertise to address shared challenges.

Together, these strategies constitute a form of teacher agency that has been largely overlooked in the literature on AI in education. While much of the existing research has focused on what AI does to teachers, this study focuses on what teachers do with AI. This shift in perspective is important not only for theoretical reasons but also for practical ones. If teachers are seen as active agents rather than passive recipients, interventions should focus on supporting and scaling their existing strategies rather than replacing them with top-down solutions.

Three Governance Dilemmas as Systemic Problems

The three dilemmas identified in this study are not failures of individual teachers but symptoms of systemic gaps in institutional support and policy. The lack of standards, time constraints, and policy gaps reflect institutional failures rather than teacher inadequacies. This distinction is critically important for designing effective interventions.

If governance dilemmas are framed as individual problems, solutions will focus on training teachers to work harder, be more careful, or develop better strategies on their own. This approach places an unfair burden on teachers and is unlikely to succeed, as it does not address the underlying structural conditions that produce the dilemmas in the first place.

If governance dilemmas are framed as systemic problems, solutions will focus on changing institutional conditions. This might include developing clear policies, providing training and support, reducing other demands on teacher time, and creating reporting and review mechanisms. This approach recognizes that teachers cannot be expected to govern GAI risks alone and that institutional support is essential for sustainable governance.

This finding aligns with recent calls in the literature for institutional approaches to AI governance. Jayasinghe et al. (2026) have proposed six institutional intervention areas to support ethical student use of GAI, including policy development, training, and technical infrastructure. Similarly, Jin et al. (2025) have documented the rapid proliferation of institutional policies for GAI in higher education globally, noting that Chinese institutions have been relatively slow to develop such policies. The findings of this study suggest that this policy gap is keenly felt by frontline teachers and that addressing it should be a priority for institutional leaders.

A Three-Level Governance Framework

Based on the findings of this study, this paper proposes a three-level governance framework that operates at the teacher, course, and institution levels. This framework is designed to be practical, scalable, and responsive to the specific challenges of ideological education.

Level 1: Teacher Level

At the teacher level, institutions should provide practical tools that reduce the burden of gatekeeping and support teachers in their existing strategies. Based on teacher recommendations in this study, these tools should include:

a. A one-page risk review checklist for common GAI risk types, organized by teaching scenario (lesson preparation, classroom instruction, interactive sessions, assessment)
b. Teaching scripts for critical questioning in class, providing language and prompts that teachers can adapt to their specific context
c. Examples of acceptable and unacceptable GAI use, developed collaboratively with teachers to ensure relevance and credibility

These tools should be developed collaboratively with teachers rather than imposed from above. The teachers in this study demonstrated considerable expertise in GAI governance through their spontaneous strategies. This expertise should be leveraged in the design of support tools.

Level 2: Course Level

At the course level, institutions should support curriculum integration that embeds AI literacy into existing courses rather than treating it as an add-on module. This integration can take multiple forms:

a. When teaching media literacy or information evaluation, instructors can incorporate examples of GAI-generated content to illustrate principles of source criticism
b. When teaching research methods, instructors can discuss how to use GAI ethically for literature review and data analysis
c. When teaching writing, instructors can develop assignments that require students to document and reflect on their use of GAI
d. The goal at the course level is not to add new content to an already crowded curriculum but to integrate AI literacy into existing learning objectives. This approach is more sustainable and more effective than standalone AI literacy modules, as it connects AI literacy to authentic disciplinary contexts.

Level 3: Institution Level

At the institution level, universities should develop clear policies and support mechanisms for GAI use. Based on the dilemmas identified in this study, these policies and mechanisms should address:

a. What constitutes acceptable and unacceptable GAI use in teaching, learning, and assessment
b. What content requires mandatory human review before classroom use
c. What training is required for teachers and what support is available
d. What reporting mechanisms exist for problematic outputs and how concerns will be addressed

Notably, these mechanisms need not be burdensome or expensive. The tools that teachers requested in this study (checklists, scripts, examples) 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.

Table 3
Three-Level Governance Framework

Level	Key Actions	Practical Tools
Teacher level	Provide operational tools for daily gatekeeping	Risk review checklist, teaching scripts, examples of acceptable/unacceptable use
Course level	Embed AI literacy into existing curriculum	Integrated modules on evaluating AI-generated content, critical questioning exercises
Institution level	Develop clear policies and support mechanisms	Use guidelines, review protocols, training programs, reporting channels

Comparison with Existing Frameworks

The proposed framework differs from existing approaches in two important ways. First, it centers teacher agency rather than teacher vulnerability. Much of the existing literature on AI in education focuses on the threats that AI poses to teacher authority and job security. While these threats are real, a focus on vulnerability can lead to defensive, reactive approaches that do little to support teachers. The proposed framework instead starts from the strategies that teachers already use, seeking to support and scale these strategies rather than replace them.

Second, the proposed framework provides concrete, operational tools rather than abstract principles. Existing calls for "strengthening value rationality" or "enhancing ethical standards" are important at a general level but offer little guidance for daily practice. The proposed framework translates these principles into specific actions and tools that teachers can use in their classrooms. This shift from principle-based warnings to practice-based governance is the study's main contribution.

The framework also aligns with recent institutional intervention models in the literature. Jayasinghe et al. (2026) propose six areas for institutional intervention: policy development, training and support, technical infrastructure, assessment design, curriculum integration, and quality assurance. The three-level framework proposed here maps onto these areas while adapting them specifically to the Chinese ideological education context.

Limitations and Future Research

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 and rich contextual understanding, it limits the generalizability of the findings to other provinces or countries with different political, cultural, or technological contexts. Future research should validate the gatekeeping strategies and governance framework across broader geographic and institutional contexts.

Second, the study relied primarily on self-reported data from interviews, which may be subject to social desirability bias or recall limitations. Teachers may have underreported risky practices or overreported their use of gatekeeping strategies. Classroom observations helped mitigate this limitation, but observational data were limited to two lessons per teacher and may not fully represent typical practice. Future research should employ more extended observation periods or diary methods to capture a fuller picture of teacher practice.

Third, the study focused on identifying gatekeeping strategies and governance dilemmas rather than testing the effectiveness of specific interventions. While the proposed framework is grounded in teacher experiences, its effectiveness has not been empirically evaluated. Future research should test the framework through intervention studies, using randomized controlled trials or quasi-experimental designs to compare outcomes for teachers who receive the checklist, scripts, and other tools versus those who do not.

Fourth, the rapid pace of GAI development means that specific risks and governance strategies may change as new models and applications emerge. The findings of this study should be treated as a snapshot of current practice rather than a fixed set of conclusions.

Longitudinal research is needed to track how gatekeeping strategies evolve as GAI technologies develop and as teachers and students gain more experience with these tools.
Fifth, this study focused exclusively on teacher perspectives. Student perspectives on GAI governance were not directly investigated. Future research should include student voices to understand how students experience teacher gatekeeping, what support they need for ethical GAI use, and how they perceive the boundaries that teachers set.

Conclusion

This study addressed the gap between risk identification and risk governance in research on Generative Artificial Intelligence in Chinese university ideological and political courses. Drawing on interviews with 20 teachers and classroom observations across five universities, the study identified four gatekeeping strategies that teachers spontaneously develop (AI output review, critical questioning, boundary setting, and peer consultation) and three governance dilemmas they face (lack of standards, time constraints, and institutional policy gaps). These findings demonstrate that teachers are active agents who develop practical approaches to GAI governance, challenging the dominant narrative of teacher vulnerability or replacement. The dilemmas are not failures of individual teachers but symptoms of systemic gaps requiring institutional action.

Based on these findings, the paper proposed a three-level governance framework operating at the teacher, course, and institution levels. The study makes two main contributions. Theoretically, it reconceptualizes the teacher as an active gatekeeper rather than a passive victim of technological change. Practically, it provides a governance framework and concrete tools that institutions can implement to support teachers. 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 effectively and equitably. This study has taken a first step toward answering that question.

References

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

Bostrom, N., & Yudkowsky, E. (2018). The ethics of artificial intelligence. In Artificial intelligence safety and security (pp. 57-69). Chapman and Hall/CRC.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

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.

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

Habermas, J. (1984). The theory of communicative action (T. McCarthy, Trans.). Beacon Press. (Original work published 1981)

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)

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

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

Lewin, K. (1947). Frontiers in group dynamics: Concept, method and reality in social science; social equilibria and social change. *Human relations*, *1*(1), 5-41.

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.

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

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)

Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.

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.

Williamson, B. (2017). Big data in education: The digital future of learning, policy and practice. SAGE Publications.

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

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). (forthcoming)

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

Zhong, H. (2025). The opportunities, challenges and countermeasures brought by the development of Chat Generative Pre-trained Transformer to person-targeted ideological and political education in colleges and universities. In Proceedings of the 2025 9th International Seminar on Education, Management and Social Sciences (ISEMSS 2025) (p. 281). Atlantis Press. https://doi.org/10.2991/978-2-38476-462-4_32