

AI Use in Higher Education Requires Risk-Graded, Actionable Guidelines

Yinpeng Dong, Zimeng Huang, Shao Chen, Alex Lamb, Jia Li, Xinyang Li, Miao Liu, Lu Mi, Yuan Yao, Xueyan Zou, Wei Xu, Andrew Chi-Chih Yao

College of AI, Tsinghua University

Abstract

The rapid adoption of AI in higher education is significantly reshaping learning, teaching, and research. Alongside gains in productivity and accessibility, unconstrained AI use introduces substantial risks, including inaccurate outputs, erosion of independent thinking, and threats to academic integrity. In response, the academic community has introduced policies governing AI use. However, the existing guidelines are largely principle-driven and provide limited operational guidance for managing AI use across academic tasks with different risks. In this position paper, we argue that **effective governance of AI in higher education requires risk-graded, actionable guidelines** rather than binary permission or prohibition. Grounded in the mission of higher education, we propose a risk-graded framework that categorizes AI use into six levels, ranging from full use to prohibition. This framework provides fine-grained, context-sensitive guidance that differentiates acceptable AI assistance across learning, teaching, and research activities. We demonstrate its practical value by establishing detailed guidelines in the field of AI and complement them with a checklist to support transparent disclosure. We call for the adoption of such guidelines across disciplines and institutions to realize the benefits of AI while safeguarding academic integrity.

Correspondence to: {dongyinpeng, weixu, andrewcyao}@tsinghua.edu.cn.

1 | Introduction

Higher education serves a central mission of educating students, advancing knowledge via research, and contributing to society (UNESCO, 1998). In recent years, the rapid evolution of AI technologies has greatly transformed educational activities (Baidoo-Anu and Ansah, 2023; Fütterer et al., 2023; Kasneci et al., 2023; Messeri and Crockett, 2024; Wang et al., 2023). Students adopt AI tools for problem solving and exploratory inquiry, instructors experiment with AI-assisted teaching and assessment; and researchers employ them for tasks such as literature review, idea generation, and academic writing. While AI provides substantial opportunities to improve learning effectiveness and research productivity, its widespread integration into higher education simultaneously raises fundamental questions about how AI tools should be used responsibly, ethically, and in alignment with core academic values.

In the absence of explicit constraints on when and how AI should be used, its rapid adoption in higher education poses significant risks (Sok and Heng, 2023; Yan et al., 2024). We mainly categorize the risks into **malfunction**, **misuse**, and **systemic risks** (Bengio et al., 2025). *Malfunction risks* stem from inherent limitations of current AI models due to their probabilistic nature, such as factual errors (Xu et al., 2024), hallucinations (Huang et al., 2025b), and biases (Gallegos et al., 2024). As a result, AI-generated content may appear fluent and authoritative while containing incorrect claims or fabricated references, thus undermining established norms of verifiability and evidence-based reasoning. For example, widespread hallucinated citations have been reported in accepted papers of NeurIPS 2025 (Shmatko et al., 2026), indicating that such failures can evade standard peer review. Beyond technical failures, *misuse risks* stem from the intentional misuse of AI for prohibited academic activities (e.g., cheating and academic fraud), where AI’s low cost, scalability, and concealability make misconduct hard to detect and undermine academic integrity (Cotton et al., 2024; Earp et al., 2025; Evangelista, 2025; Goodier, 2025; Huang et al., 2025a). Finally, *systemic risks* emerge through sustained reliance on AI, which may erode essential human capacities such as independent thinking, posing a cumulative threat to the foundations of education and research (Chan and Lee, 2023; Darvishi et al., 2024; Hao et al., 2026).

In response, universities have issued formal guidelines governing AI use (An et al., 2025; Chan, 2023; Porsdam Mann et al., 2024). These policies typically emphasize transparency, human accountability, and data privacy. Similar norms have been adopted in scholarly publishing: for instance, Springer Nature stresses author responsibility and peer-review confidentiality¹, while ICML 2026 allows AI use in writing or research, but requires author responsibility and encourages the disclosure². Despite these efforts, existing guidelines remain insufficient. First, most of them are principle-heavy but behavior-ambiguous, offering high-level expectations (e.g., “responsible use”) without delineating acceptable actions in routine academic tasks. Moreover, they often treat “AI use” as a broad category, overlooking how risks vary across different stages (e.g., brainstorming vs. writing) of various academic activities (e.g., research vs. teaching). Consequently, low-stakes uses such as language polishing are governed alongside high-stakes applications like experimental design, limiting the operational value of requirements for disclosure and verification. These underscore the need for governance approaches that move beyond the general principles and instead provide **context-sensitive** guidance for AI use in higher education.

In this paper, we argue that **AI use in higher education requires risk-graded, actionable guidelines**. Our goal is to provide practical guidance for students and researchers, particularly those at early stages of their academic training, on responsible AI use. Considering the core goals of varying contexts of higher education, e.g., cultivating independent thinking in teaching and advancing knowledge in research, we adopt a risk-centered approach that explicitly accounts for context-specific risks of AI. Specifically, we propose a risk-graded framework categorizing AI use into six levels—*full use, encouraged use, moderate use, cautious use, restricted use, and prohibited use*. Rather than treating AI use as a binary choice, our framework enables risk-aware, fine-grained assistance across academic workflows, offering flexible, context-sensitive, and operational guidance over principle-based policies.

As a case study, we apply this framework to teaching and research activities in the field of AI and develop detailed, context-sensitive guidelines specifying appropriate AI use levels across common academic scenarios. For example, we encourage AI use for literature search but data fabrication is strictly prohibited. These guidelines were formulated by a committee of over ten faculty members from a leading AI institution and were informed by consultations with more than fifty AI researchers and doctoral students. We further complement the guidelines with a practical checklist (Appendix C) to support transparent disclosure. While our case study only focuses on AI, we encourage the adoption of our framework across fields and institutions to facilitate using AI more responsibly and effectively.

2 | Opportunities and Risks of AI in Higher Education

2.1 | Widespread Use of AI

AI adoption in higher education has accelerated, driven by stronger AI capabilities and proactive institutional uptake. Global surveys report that AI had become routine for 86% of university students by 2024 (Digital Education Council, 2024) and nearly universal among undergraduates by 2025, with 92% reporting some use (Freeman, 2025). Faculty adoption has also risen, with over 60% of educators using AI in teaching by 2025 (Digital Education Council, 2025). While AI can personalize learning and improve efficiency, misuse and overreliance remain concerns: 8% of students report submitting unmodified AI-generated text (Pearson, 2025), and excessive dependence may undermine cognitive engagement and critical thinking (Kosmyna et al., 2025).

AI is also reshaping research by supporting hypothesis generation, experimental design, and data analysis (Wang et al., 2023). Evidence from arXiv suggests that LLM-assisted writing has already affected scientific production: from January 2022 to July 2024, LLM-using authors published 36.2% more preprints per month than comparable non-adopters (Kusumegi et al., 2025). AI tools also lower language barriers and aid literature discovery. Yet scaling AI-assisted research also scales concerns: AI is expected to become routine in writing, peer review, and grant drafting, while potentially increasing low-quality submissions and AI-assisted fakes (Conroy, 2023a). Undisclosed AI-assisted manuscripts can evade peer review, motivating disclosure and practical checks (Conroy, 2023b). Besides, LLM-assisted ICLR 2024 submissions showed increased writing complexity correlated with lower reviewer scores and acceptance probability (Kusumegi et al., 2025), and hallucinated references in accepted NeurIPS 2025 papers reveal peer review’s limits in detecting AI errors (Shmatko et al., 2026).

¹<https://www.springernature.com/gp/policies/editorial-policies>

²<https://icml.cc/Conferences/2026/CallForPapers>

2.2 | Risk Taxonomy

Given the broad adoption of AI and the growing concerns, it is essential to first identify and characterize the associated risks. Following [Bengio et al. \(2025\)](#), we distinguish three categories: *malfunction risks*, *misuse risks*, and *systemic risks*. In educational and research contexts, the risks can **(1) undermine academic integrity and scientific rigor**, and **(2) erode the goals of education and research**.

Malfunction Risks. Malfunction risks arise from limitations in AI reliability, including hallucinations and inconsistent performance. Driven by probabilistic objectives, AI models often prioritize fluency over factual accuracy ([Zhou et al., 2024](#)). The failures include factual errors and fabricated citations ([Huang et al., 2025b](#); [Ji et al., 2023](#)), unsafe code generation ([Lian et al., 2025](#)), logical inconsistencies in reasoning chains ([Lanham et al., 2023](#); [Shojaee et al., 2025](#)), and social or epistemic biases ([Gallegos et al., 2024](#)). Moreover, recent work identifies deceptive or misaligned behaviors in frontier models ([Greenblatt et al., 2024](#); [Huang et al., 2025c](#)). These failures primarily **threaten scientific rigor**, though their impact can be mitigated through expert verification and responsible human oversight.

Misuse Risks. Misuse risks arise when users intentionally exploit AI to bypass learning objectives or violate academic norms. Examples include AI-generated student submissions and AI-assisted cheating ([Buck, 2025](#); [Reiter et al., 2025](#)), data fabrication or manipulation in research ([Májovský et al., 2023](#)), and the delegation of high-stakes academic judgments to AI without oversight ([Mammadova et al., 2025](#)). By lowering the cost of misconduct, AI enables academic dishonesty at scale ([Earp et al., 2025](#); [Evangelista, 2025](#)), while simultaneously becoming increasingly difficult to detect ([Goodier, 2025](#); [Huang et al., 2025a](#)). Beyond integrity violations, such misuse **undermines the core function of education** by invalidating assessment and eroding the credibility of academic credentials.

Systemic Risks. Systemic risks are cumulative, long-term consequences that extend beyond isolated failures or misconduct. In education, the central concern is the erosion of human cognitive capabilities: sustained AI reliance may weaken critical thinking, independent inquiry, analytical reasoning, and foundational skills such as problem decomposition and abstraction ([Chan and Lee, 2023](#); [Chen and Gong, 2025](#); [Cotton et al., 2024](#); [Darvishi et al., 2024](#); [Tian and Zhang, 2025](#); [Zhai et al., 2024](#)). In research, systemic risks threaten scientific innovation: although AI can increase productivity, large-scale evidence suggests it may narrow intellectual diversity, reduce cross-disciplinary exploration, and homogenize research ([Hao et al., 2026](#)). These cumulative effects **compromise the core goals of education and research**, motivating governance beyond ad hoc or binary restrictions.

3 | Existing AI Use Guidelines Are Insufficient

To mitigate these risks and promote responsible AI use in higher education, the academic community has begun developing formal guidelines ([An et al., 2025](#); [Chan, 2023](#); [Porsdam Mann et al., 2024](#)). We next discuss the guidelines and their limitations.

3.1 | Existing Guidelines for Using AI

Many universities have issued AI usage policies or guidance documents³⁴⁵, emphasizing responsible and ethical use in academic work. To identify common patterns, we surveyed policies from over 20 prominent universities across regions and disciplines. Despite institutional variation, these policies converge on several principles: (1) academic integrity; (2) transparency and, where appropriate, disclosure of AI assistance; (3) instructor- or course-level autonomy over acceptable use; (4) AI as an aid rather than a substitute for student work; (5) student or faculty responsibility for verifying AI-generated outputs; and (6) data privacy and confidentiality, including limits on sharing sensitive materials with third-party tools.

Similar guidelines have emerged in scholarly publishing and major conferences. Leading publishers now address AI-assisted research and writing: Springer Nature, for instance, emphasizes author accountability for all submitted content and peer-review confidentiality when AI tools are used¹. Major AI conferences have likewise introduced AI-use policies for submission and review. ICML 2026 restricts large language models in core evaluative judgments while holding reviewers responsible for final content²,

³<https://communitystandards.stanford.edu/generative-ai-policy-guidance>

⁴<https://provost.harvard.edu/guidelines-using-chatgpt-and-other-generative-ai-tools-harvard>

⁵<https://libguides.princeton.edu/generativeAI/disclosure>

with comparable constraints at venues such as NeurIPS⁶. Across these settings, common requirements include human responsibility, no AI authorship, avoidance of plagiarism or fabrication, and disclosure of substantive AI assistance.

3.2 | Limitations of Existing Guidelines

Despite previous education-oriented or research-driven efforts, our community still lacks a clear and systematic articulation of guidelines that effectively address the realities of AI use in higher education. Specifically, existing guidelines fall short in the following key perspectives.

First, most existing policies, especially those issued at the university level, are developed primarily for administrative purposes and are therefore principle-heavy while remaining operationally underspecified. For example, a frequent guideline format states high level expectations using terms such as “responsible use” or “transparency”, without clearly delineating acceptable versus unacceptable forms of AI assistance in common educational and research activities.

Second, whether AI use is appropriate depends strongly on the specific scenario, yet such context sensitivity is largely missing from prior efforts. Notably, existing guidelines rarely provide a fine-grained characterization of different modes of AI use, such as brainstorming, analytical support, or grammar related assistance, which makes them difficult for the target audience to interpret and follow.

Third, AI use in educational settings should not be framed as a binary decision problem, but instead calls for a categorical, risk-level based reference conditioned on specific scenarios and user profiles. For example, AI assistance in language polishing poses qualitatively lower risks than its use in peer review, while assistance during examinations should be fully prohibited. However, such risk-level distinctions are largely omitted from existing guidelines.

Therefore, our position is timely in filling gaps in previous guidelines and aims to mitigate risks that may undermine learning processes and academic integrity, while still harnessing AI to support more productive educational practices. Although AI use guidelines have become more detailed this year, a gap remains between current guidance and our proposal. We provide a more detailed discussion with concrete examples in Appendix D.

4 | Proposal: Establish Risk-Graded, Actionable AI Guidelines

In this section, we propose a framework for establishing risk-graded, actionable guidelines. We first present the guiding principles and then develop a six-level risk-graded framework (Table 1).

4.1 | General Principles

Our proposal is built upon three core principles that reveal our essential understanding and position on AI use. We expect that this guideline reflects both a positive vision for the developmental potential of AI technology and the core values that should be upheld in talent cultivation and academic research.

We encourage the proactive application of AI in educational practice. While harnessing its productivity gains, we argue that data privacy, critical thinking, academic integrity, and ethical responsibility must remain central to all academic endeavors. To this end, we propose the following principles as general guidelines for using AI technologies.

Principle 1. Accountability and oversight: Emphasizing human as the primary subject. Given that the fundamental mission of higher education is research innovation and talent development, we encourage faculty and students to leverage AI to expand intellectual horizons and enhance efficiency, provided that all usage strictly adheres to data privacy and security regulations. However, the process of using AI should prioritize the cultivation and growth of individual academic capabilities. We argue that faculty and students should recognize the human as the primary subject in education and research, and bear full responsibility for all academic outputs. Thus, humans should be fully responsible for the scientific rigor, compliance, and originality of any AI-generated or AI-assisted content in their work.

Principle 2. Transparent disclosure: Fostering a trustworthy academic environment. Disclosure of AI use is a fundamental requirement for accountability and is essential for maintaining academic integrity

⁶<https://neurips.cc/Conferences/2025/LLM>

Table 1: A six-level risk-graded framework for AI use in higher education.

Level	Rating	Recommended Use	Dominant Risks	Description of Risks
Full Use	★★★★★	AI should be fully leveraged to improve efficiency and reduce repetitive, non-developmental work.	Negligible	–
Encouraged Use	★★★★☆	AI use is encouraged to broaden perspectives, stimulate ideas, and accelerate information access.	Malfunction Risks	Factual errors; hallucinations; unreliable reasoning; AI deception; bias.
Moderate Use	★★★☆☆	AI may assist under explicit human-in-the-loop constraints; it should not be used to generate core judgments or determine research direction.	Systemic Risks	Over-reliance without critical engagement; reduced competency development; path dependence.
Cautious Use	★★☆☆☆	AI use in critical stages should be carefully bounded and sequenced; direct generation of original ideas or solution paths should be avoided.	Systemic Risks	Substitution of deep thinking; reduced originality and intellectual diversity; erosion of long-term innovative capacity.
Restricted Use	★☆☆☆☆	AI use is generally discouraged and allowed only under strict conditions that ensure containment and accountability.	Systemic Risks	Loss of independent judgment and capacity for originality.
Prohibited Use	☆☆☆☆☆	AI use is prohibited in principle.	Misuse Risks & Systemic Risks	Academic misconduct; privacy and data security violations; substitution of core learning or innovation.

and credibility. To build an open, trustworthy, and traceable academic environment, faculty and students should provide clear, specific, and necessary disclosure of AI use in teaching, research, and related academic outputs, based on actual circumstances. Disclosure should generally include the scope, manner, and role of AI use in forming the output. Relevant scenarios include course assignments or projects, dissertations, theses, academic papers, and other academic outputs. Specific disclosure requirements and a reference checklist are provided in Appendix C.

Principle 3. Dynamic evolution: Establishing an open and evolving guidance framework. AI technology is rapidly developing, with its capabilities, application scenarios, and risks constantly evolving. As capabilities advance, AI may become more feasible, reliable, and useful in teaching and research, expanding its usable boundaries and application scenarios. At the same time, the likelihood and complexity of risks will also change. Therefore, the guidelines should be continuously evaluated, adjusted, and refined based on technological progress, practical experience, and community feedback.

4.2 | A Risk-Graded Framework

Based on the core goals of higher education and the philosophy of talent cultivation, we propose a risk-graded framework that classifies the extent of AI use into six levels: *full use*, *encouraged use*, *moderate use*, *cautious use*, *restricted use*, and *prohibited use*. According to our risk taxonomy in Section 2.2, risks fall into three categories. The framework takes into account the nature of the risks, and the severity of each risk determines the recommended level of AI use. Different levels thus apply to specific scenarios to intuitively present the recommended use for the corresponding risks. However, as stated in Principle 1 in Section 4.1, it is humans who make the final grading judgment in specific contexts. In the following section, in line with the framework’s conceptual approach and methodology, a case study will demonstrate how it works in practice.

Level 5: Full Use (★★★★★). This applies to low-stakes, easily verifiable, and largely mechanical activities where AI can reduce repetitive labor without substituting for core intellectual work. When used appropriately, AI can significantly enhance efficiency while still supporting fundamental educational objectives. Therefore, in such contexts, we not only allow but actively encourage students to leverage AI as a necessary tool to achieve their learning and research training goals more effectively.

Level 4: Encouraged Use (★★★★☆). At this level, AI use is encouraged to broaden perspectives, support ideation, and speed up information acquisition. The dominant risks arise from AI malfunctions (e.g., factual errors and inaccuracies, hallucinations, and logical reasoning errors), which can mislead users if accepted uncritically. Risk control is needed through human verification and by treating AI outputs as suggestions rather than the truths.

Level 3: Moderate Use (★★★☆☆). This level permits AI assistance for analysis and sensemaking under explicit human-in-the-loop constraints, while prohibiting AI from producing the core judgment, conclusion, or research path. Unlike the previous two levels, starting from this level, systemic risks become salient: users may unintentionally delegate responsibility and lose competency development at the same time, or allow AI to shape decisions without independent judgment or critical thinking.

Level 2: Cautious Use (★★☆☆☆). This is reserved for critical stages where the academic objective is precisely to develop deep understanding, original thinking, or independent research formulation, such as defining research questions or proposing solution strategies. Compared with Level 3, both the probability and severity of systemic risks are further escalated in these contexts: even correct AI outputs may gradually substitute for cognitive effort, narrowing intellectual exploration and weakening originality. Over the long term, academic innovation will be undermined. Thus, AI use must be carefully sequenced and bounded. For instance, using AI only after an initial human attempt or only for limited critique rather than generation.

Level 1: Restricted Use (★☆☆☆☆). Compared with *cautious use*, this level imposes stricter limits because the associated risk becomes inevitable without strict constraints. It applies specifically to mitigate the risk of long-term dependency, where the loss of independent judgment and capacity for originality becomes a primary concern. The purpose is to prevent AI from becoming a crutch that erodes independent judgment. Thus, AI may only be used under strict controls that mitigate the risk.

Level 0: Prohibited Use (☆☆☆☆☆). This applies to activities whose epistemic or institutional role is fundamentally incompatible with AI assistance. The misuse risks (e.g., privacy breaches, academic misconduct) are severe. Therefore, AI use is prohibited in principle to safeguard academic integrity, procedural fairness, and security obligations. Additionally, activities that are core to learning and innovation are irreplaceable by AI, and their exposure to systemic risks must be explicitly prevented.

5 | Case Study

To illustrate how our risk-graded framework enables actionable governance, we establish a set of detailed guidelines in the field of AI covering both *Teaching* and *Research*. We focus here on the research part to demonstrate its application. The guidelines were first drafted by a committee of over ten faculty members from a world-class AI institution. More than fifty AI researchers and doctoral students then voted on appropriate AI-use levels across academic tasks and provided feedback. The committee revised the guidelines over several rounds; for disputed items, decisions were made through open deliberation grounded in the framework's risk criteria and majority voting.

This section elects to present **AI Guidelines for AI Research**, as shown in Figure 1, which align risk levels with the research lifecycle (Luo et al., 2025); then introduces a **Risk-Graded AI Compliance Checklist** to support actionable and transparent disclosure; and concludes with a **Call to Action**, urging broader adoption and adaptation of these instruments across disciplines. The detailed guidelines for research and teaching are provided in Appendix A and Appendix B.

5.1 | AI Guidelines for AI Research

Stage 1: Literature Review. The literature review stage establishes the knowledge base for subsequent research. While AI can significantly improve efficiency in information retrieval and organization, its use must be constrained to avoid substituting AI outputs for researchers' own reading and understanding. Accordingly, in our guidelines, *Full Use (L5)* includes **Literature Management**, allowing AI to organize bibliographic entries and citation-related notes efficiently. *Encouraged Use (L4)* includes **Literature Search** and **Literature Filtering** to broaden information access. *Moderate Use (L3)* applies to **Translation Assistance** and **Conceptual Understanding**, which require verification by humans. Stricter constraints apply to cognitively demanding tasks. *Cautious Use (L2)* covers **Paper Summarization** and **Paper Comparison**, which need deep thinking of humans. Using AI as a substitute for **Deep Reading** is



Figure 1: AI guidelines for AI research. The x-axis denotes the sequential stages of research, while the y-axis represents the AI use levels. The numerical values within each task block correspond to the star ratings (L0–L5) indicated on the vertical axis, representing the recommended intensity of AI involvement and the required degree of human oversight. The “Actionable Cases” panel highlights concrete examples of aligned practices (“Do’s”) versus critical violations (“Don’ts”), demonstrating how to distinguish between efficiency-enhancing assistance and integrity-compromising reliance across different phases.

regarded as *Prohibited Use (L0)*, as direct engagement with foundational literature is essential for sound epistemic judgment.

Stage 2: Research Ideation. The research ideation stage carries significant systemic risks of epistemic homogenization, requiring a careful balance between efficiency and creativity. *Encouraged Use (L4)* comprises **Novelty Check** and **Limitation Analysis**, which facilitate the quick positioning of work against the status quo. Once a plan exists, *Moderate Use (L3)* includes **Proposal Refinement** for refining structural details. However, stricter constraints are necessary for tasks involving core creativity. *Cautious Use (L2)* encompasses **Research Formulation**, **Research Design**, and **Method Selection**. If AI dictates the research problem or technical path, the work risks converging on the statistically probable mean of the training data. Furthermore, *Restricted Use (L1)* applies to **Research Assessment**, as AI should not be the sole arbiter of an idea’s merit.

Stage 3: Theoretical Analysis. In theoretical work, AI should act as a surveyor and checker, not a prover. *Encouraged Use (L4)* permits **Theory Review** to broaden context. *Moderate Use (L3)* covers **Proof Assistance** and **Proof Check**, where AI may suggest lemmas or flag inconsistencies under strict human supervision. *Cautious Use (L2)* applies to **Problem Formalization** and **Hypothesis Generation**, requiring alignment with researcher intent. *Prohibited Use (L0)* includes **Unverified Derivation** and **Proof Fabrication**, e.g., unreproduced AI-generated proofs or masked logical gaps.

Stage 4: Experimentation. This phase distinguishes between routine engineering and scientific inquiry. *Full Use (L5)* includes **Boilerplate Code** (e.g., wrappers, visualization) to effectively reduce labor. *Encouraged Use (L4)* covers **Data Preprocessing** and **Documentation** to enhance reproducibility. However, scientific decisions require human agency. *Moderate Use (L3)* applies to **Experimental Design**, assisting but not replacing the researcher’s plan. *Cautious Use (L2)* encompasses **Data Analysis**, **Implementation Assistance**, and **Pipeline Design**. Researchers who rely on AI for core innovations or statistical testing may misunderstand the methodology or misinterpret significance. **Data Fabrication** remains *Prohibited Use (L0)*.

Stage 5: Scholarly Writing. We grade writing tasks by semantic load and implications for academic

equity (Liang et al., 2023). *Full Use (L5)* covers **Polishing & Formatting**, supporting non-native speakers, while *Moderate Use (L3)* covers **Structural Outlining**. Content generation requires stricter limits: *Cautious Use (L2)* applies to **Related Work Writing** due to systemic risks, and *Restricted Use (L1)* to **Method Writing**, where bypassing authorship can erode rigor. *Prohibited Use (L0)* covers **Core Contribution Writing** and **Reference Generation**, preventing AI-authored claims and fabricated citations.

Stage 6: Peer Review. Peer review relies on trusted expert judgment. *Moderate Use (L3)* includes **Reading Assistant**, allowing reviewers to summarize long papers or check formatting. However, *Prohibited Use (L0)* applies to **Review Generation**. The assessment of novelty and validity is the reviewer's responsibility and cannot be delegated to AI.

Stage 7: Rebuttal. During the rebuttal phase, authors find *Encouraged Use (L4)* in **Rebuttal Organization** to help organize feedback from multiple reviewers. Conversely, *Prohibited Use (L0)* applies to **Rebuttal Generation**. The defense of scientific work, including additional analysis and argumentation, must be authored by the researchers themselves to maintain integrity.

Limitations. 1) Some level assignments in our risk-graded guidelines remain debatable. Consultations with AI researchers and students revealed disagreement on the appropriate risk level for certain tasks, reflecting differences across contexts and experience levels. Thus, the guidelines are advisory rather than mandatory, and should be revised as AI capabilities and academic needs evolve. 2) Our concrete guidelines focus on AI research and teaching as an illustrative case. However, the framework is designed to generalize across disciplines because it is organized around the shared mission of higher education. Institutions can adapt it by identifying discipline-specific tasks, risks, and use levels, as discussed in Appendix D. 3) The guidelines improve transparency and clarify expectations, but cannot prevent malicious misuse. Intentional violations may remain undisclosed, so this framework does not replace auditing or enforcement. Instead, it helps reduce misuse caused by ambiguity about acceptable AI use, as discussed in Appendix D.

5.2 | Risk-Graded AI Compliance Checklist

To operationalize risk-graded governance in everyday research, we introduce a **Risk-Graded AI Compliance Checklist** (Appendix C). While guidelines articulate normative principles, this checklist provides a concrete mechanism for compliance. The checklist is intentionally concise and execution-oriented. It decomposes the research lifecycle (Stages 1–5) into stage-specific, binary question pairs (4–5 per stage) that authors can quickly verify (e.g., “Did you use AI for literature search or filtering?”). This design enables efficient self-assessment while preserving human responsibility for high-risk research decisions. *As AI was also used in this work, the responses in the checklist provided in Appendix C explicitly disclose how AI tools were employed in the preparation of this paper.*

We propose integrating the checklist into conference and journal submission systems as a required disclosure. By attesting to actions such as AI-proof validation and data-fabrication checks, authors take primary responsibility for research integrity, shifting governance from retrospective policing to proactive accountability. The checklist also gives reviewers and the broader community with a clear signal of AI involvement and human oversight, helping sustain trust in academic outputs.

5.3 | Call to Action

The rapid integration of AI into higher education calls for a shift from binary “ban-or-allow” policies to risk-graded governance. We propose three actionable recommendations:

- 1. Adopt and adapt risk-graded frameworks.** We urge institutions to move beyond one-size-fits-all restrictions and adopt risk-graded frameworks tailored to disciplinary goals and ethical contexts. Explicit risk boundaries enable responsible AI use while preserving academic rigor.
- 2. Operationalize transparency with standardized checklists.** Effective policy requires actionable mechanisms. We call on conferences, publishers, and funding bodies to integrate standardized AI compliance checklists into submission and review workflows. Such disclosure institutionalizes accountability and strengthens the integrity of peer review.
- 3. Enable continuous review and evolution.** As AI capabilities evolve, governance must remain adaptive. We recommend establishing standing committees within universities and professional societies to regularly assess and update AI guidelines through community-driven dialogue.

We have begun applying the framework and guidelines in teaching and research activities at AI-related institutions in our university. Students and researchers are informed of the guidelines and are expected to use them as practical references for responsible AI use. We view this as an initial step toward real-world validation, and will collect surveys, interviews, and checklist responses to evaluate whether the guidelines help users use AI more responsibly.

6 | Alternative Views

While we advocate for concrete, risk-graded, and actionable guidelines for AI usage, several alternative views have been advocated for. One view advocates near-total prohibition of AI in education. For example, [Berg \(2025\)](#) argues against AI use even in seemingly benign tasks such as summarization, contending that no component of cognitive understanding is dispensable. Others have supported guidelines for AI, but have framed these guidelines in terms of foundational ethical principles, without introducing risk-grading and recommendations for specific types of AI usages. Another alternative position suggests that AI usage should vary on a case-by-case basis, considering the particular conditions which apply to different students or institutions.

6.1 | AI Use Should Be Strictly Prohibited

In early 2023, some universities (e.g., Oxbridge) banned the use of ChatGPT due to concerns about cheating, plagiarism, and unauthorized assistance ([Xiao et al., 2023](#)). This gives rise to a stance that strictly prohibiting the use of AI may be the most reliable way to maintain academic integrity ([Berg, 2025](#)). Proponents argue that inconsistent or permissive policies may create exploitable loopholes that erode academic standards. Additional concerns include the environmental cost of AI's energy consumption, potential violations of student privacy through third-party data access, and uncertainty about the long-term effects of AI use on learning outcomes ([de Fine Licht, 2024](#)).

Response: Despite recognizing these legitimate concerns, we contend that strict prohibition can unnecessarily sacrifice legitimate, low-risk uses of AI, such as repetitive and mechanical activities. As our analysis shows, the nature and impact of risks vary across academic tasks, and differentiated, risk-graded guidance can better safeguard integrity while preserving the benefits of using AI.

6.2 | Principle-Based Guidelines Are Sufficient

Another view is that high-level, principle-based guidelines, such as transparency, human accountability, and academic integrity, are sufficient for governing AI use, as common practice ([Nguyen et al., 2023](#); [Adams et al., 2023](#); [Nguyen, 2025](#)). Proponents argue that prescriptive rules risk becoming obsolete as AI evolves, and that flexible principles better preserve academic autonomy across fields and institutions.

Response: While principles are necessary, we argue that they are insufficient. In practice, broadly framed guidance often leaves substantial ambiguity in everyday academic decision-making, forcing individuals to interpret abstract norms at each task. This burden falls disproportionately on early-stage students and leads to inconsistent application, confusion, and unintentional misuse. Our risk-graded framework is intended not to replace them, but to operationalize them in a manner that remains adaptable as AI evolves.

6.3 | AI Use Should Depend on Individual Judgment

The third view argues that decisions about appropriate AI use should depend on individual judgment. Based on this view, centralized guidelines may inadequately capture disciplinary norms, pedagogical goals, or contextual nuances, and may constrain legitimate exploration with new tools ([Truong et al., 2025](#); [Farinosi and Melchior, 2025](#)).

Response: We share the concern that rigid, one-size-fits-all rules can be misaligned with local contexts. Nevertheless, reliance on individual judgment alone can exacerbate inequities and uncertainty, particularly for students navigating inconsistent expectations across courses, advisors, or institutions. Our guidelines are advisory rather than mandatory. They are designed to support individual judgment by providing actionable items for using AI. By making implicit expectations explicit, risk-graded guidelines can reduce ambiguity while still allowing individuals and institutions to adapt specific implementations to their needs.

7 | Conclusion

In this paper, we argued that AI use in higher education requires risk-graded, actionable guidelines. We proposed a six-level framework that differentiates AI use according to task characteristics and associated risks. We demonstrated the practical utility of this approach through a detailed instantiation in the context of AI education and research, providing task-level guidelines across teaching and research workflows, as well as a disclosure checklist to support transparency. While our concrete examples focus on AI as a discipline, the underlying framework is designed to be extensible and adaptable to other fields and institutional contexts. We view this work as a starting point rather than a final prescription. As AI technologies evolve rapidly, so too must the norms and guidelines that govern their use. Risk-graded, advisory frameworks offer a pragmatic path forward—one that balances flexibility with responsibility, and efficiency with the preservation of core academic values.

References

- Catherine Adams, Patti Pente, Gillian Lerner, and Geoffrey Rockwell. Ethical principles for artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence*, 4:100131, 2023.
- Yunjo An, Ji Hyun Yu, and Shadarra James. Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. *International Journal of Educational Technology in Higher Education*, 22(1):10, 2025.
- David Baidoo-Anu and Leticia Owusu Ansah. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1):52–62, 2023.
- Yoshua Bengio, Sören Mindermann, Daniel Privitera, Tamay Besiroglu, Rishi Bommasani, Stephen Casper, Yejin Choi, Philip Fox, Ben Garfinkel, Danielle Goldfarb, et al. International AI safety report. *arXiv preprint arXiv:2501.17805*, 2025.
- Anastasia Berg. Why even basic A.I. use is so bad for students. 2025. URL <https://www.nytimes.com/2025/10/29/opinion/ai-students-thinking-school-reading.html>.
- Daniel Buck. From crib sheets to AI cheats, everyone's doing it. *Education Next*, 25(1):1–1, 2025.
- Cecilia Ka Yuk Chan. A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education*, 20(1):38, 2023.
- Cecilia Ka Yuk Chan and Katherine KW Lee. The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and millennial generation teachers? *Smart Learning Environments*, 10(1):60, 2023.
- Chen Chen and Yang (Frank) Gong. The role of AI-assisted learning in academic writing: A mixed-methods study on chinese as a second language students. *Education Sciences*, 15(2), 2025.
- Gemma Conroy. How ChatGPT and other AI tools could disrupt scientific publishing. *Nature*, 622(7982):234–236, 2023a.
- Gemma Conroy. Scientific sleuths spot dishonest ChatGPT use in papers. *Nature*, 2023b.
- Debby RE Cotton, Peter A Cotton, and J Reuben Shipway. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2):228–239, 2024.
- Ali Darvishi, Hassan Khosravi, Shazia Sadiq, Dragan Gašević, and George Siemens. Impact of AI assistance on student agency. *Computers & Education*, 210:104967, 2024.
- Karl de Fine Licht. Generative artificial intelligence in higher education: Why the 'banning approach' to student use is sometimes morally justified. *Philosophy & Technology*, 37, 2024.
- Digital Education Council. Global AI faculty survey 2024. <https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-student-survey-2024>, August 2024. A survey provides insights into student perceptions around AI in higher education.
- Digital Education Council. Global AI faculty survey 2025. <https://www.digitaleducationcouncil.com/post/digital-education-council-global-ai-faculty-survey>, January 2025. A survey provides insights into faculty use and perceptions of AI in higher education.

- Brian D Earp, Haotian Yuan, Julian Koplin, and Sebastian Porsdam Mann. LLM use in scholarly writing poses a provenance problem. *Nature Machine Intelligence*, 7:1889–1890, 2025.
- Edmund De Leon Evangelista. Ensuring academic integrity in the age of chatgpt: Rethinking exam design, assessment strategies, and ethical AI policies in higher education. *Contemporary Educational Technology*, 17(1):ep559, 2025.
- Manuela Farinosi and Claudio Melchior. To adopt or to ban? Student perceptions and use of generative AI in higher education. *Humanities and Social Sciences Communications*, 12(1):1–14, 2025.
- Josh Freeman. Student generative AI survey 2025. *Higher Education Policy Institute: London, UK*, 2025.
- Tim Fütterer, Christian Fischer, Anastasiia Alekseeva, Xiaobin Chen, Tamara Tate, Mark Warschauer, and Peter Gerjets. ChatGPT in education: Global reactions to AI innovations. *Scientific Reports*, 13(1):15310, 2023.
- Isabel O Gallegos, Ryan A Rossi, Joe Barrow, Md Mehrab Tanjim, Sungchul Kim, Franck Dernoncourt, Tong Yu, Ruiyi Zhang, and Nesreen K Ahmed. Bias and fairness in large language models: A survey. *Computational Linguistics*, 50(3):1097–1179, 2024.
- Michael Goodier. Revealed: Thousands of UK university students caught cheating using AI. *The Guardian*, 15, 2025.
- Ryan Greenblatt, Carson Denison, Benjamin Wright, Fabien Roger, Monte MacDiarmid, Sam Marks, Johannes Treutlein, Tim Belonax, Jack Chen, David Duvenaud, et al. Alignment faking in large language models. *arXiv preprint arXiv:2412.14093*, 2024.
- Qianyue Hao, Fengli Xu, Yong Li, and James Evans. Artificial intelligence tools expand scientists' impact but contract science's focus. *Nature*, pages 1–7, 2026.
- Dongpeng Huang, Nicole Hash, James J Cummings, and Kelsey Prena. Academic cheating with generative AI: Exploring a moral extension of the theory of planned behavior. *Computers and Education: Artificial Intelligence*, page 100424, 2025a.
- Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):1–55, 2025b.
- Yao Huang, Yitong Sun, Yichi Zhang, Ruochen Zhang, Yinpeng Dong, and Xingxing Wei. DeceptionBench: A comprehensive benchmark for AI deception behaviors in real-world scenarios. In *Neural Information Processing Systems*, 2025c.
- Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38, 2023.
- Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274, 2023.
- Nataliya Kosmyna, Eugene Hauptmann, Ye Tong Yuan, Jessica Situ, Xian-Hao Liao, Ashly Vivian Beresnitzky, Iris Braunstein, and Pattie Maes. Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv preprint arXiv:2506.08872*, 4, 2025.
- Keigo Kusumegi, Xinyu Yang, Paul Ginsparg, Mathijs de Vaan, Toby Stuart, and Yian Yin. Scientific production in the era of large language models. *Science*, 390(6779):1240–1243, 2025.
- Tamera Lanham, Anna Chen, Ansh Radhakrishnan, Benoit Steiner, Carson Denison, Danny Hernandez, Dustin Li, Esin Durmus, Evan Hubinger, Jackson Kernion, et al. Measuring faithfulness in chain-of-thought reasoning. *arXiv preprint arXiv:2307.13702*, 2023.
- Keke Lian, Bin Wang, Lei Zhang, Libo Chen, Junjie Wang, Ziming Zhao, Yujiu Yang, Miaoqian Lin, Haotong Duan, Haoran Zhao, et al. A.S.E: A repository-level benchmark for evaluating security in AI-generated code. *arXiv preprint arXiv:2508.18106*, 2025.
- Weixin Liang, Mert Yuksekgonul, Yining Mao, Eric Wu, and James Zou. GPT detectors are biased against non-native English writers. *Patterns*, 4(7), 2023.
- Ziming Luo, Zonglin Yang, Zexin Xu, Wei Yang, and Xinya Du. LLM4SR: A survey on large language models for scientific research. *arXiv preprint arXiv:2501.04306*, 2025.
- Martin Májovský, Martin Černý, Matěj Kasal, Martin Komarc, David Netuka, et al. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. *Journal of Medical Internet Research*, 25(1):e46924, 2023.
- Ilhama Mammadova, Fatime Ismayilli, Elnaz Aliyeva, and Narmin Mammadova. The role of artificial intelligence in shaping assessment practices in higher education. *Educational Process: International Journal*, 18:e2025439, 2025.

- Lisa Messeri and Molly J Crockett. Artificial intelligence and illusions of understanding in scientific research. *Nature*, 627(8002):49–58, 2024.
- Andy Nguyen, Ha Ngan Ngo, Yvonne Hong, Belle Dang, and Bich-Phuong Thi Nguyen. Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28(4):4221–4241, 2023.
- Khoa Viet Nguyen. The use of generative ai tools in higher education: Ethical and pedagogical principles. *Journal of Academic Ethics*, pages 1–21, 2025.
- Helen Pearson. Universities are embracing AI: Will students get smarter or stop thinking? *Nature*, 646(8086):788–791, 2025.
- Sebastian Porsdam Mann, Anuraag A Vazirani, Mateo Aboy, Brian D Earp, Timo Minssen, I Glenn Cohen, and Julian Savulescu. Guidelines for ethical use and acknowledgement of large language models in academic writing. *Nature Machine Intelligence*, 6(11):1272–1274, 2024.
- Leonhard Reiter, Moritz Jörling, Christoph Fuchs, and Robert Böhm. Student (mis)use of generative AI tools for university-related tasks. *International Journal of Human–Computer Interaction*, 41(19):12390–12403, 2025.
- Nazar Shmatko, Alex Adam, and Paul Esau. GPTZero finds 100 new hallucinations in NeurIPS 2025 accepted papers. <https://gptzero.me/news/neurips/>, January 2026.
- Parshin Shojaee, Iman Mirzadeh, Keivan Alizadeh, Maxwell Horton, Samy Bengio, and Mehrdad Farajtabar. The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. *arXiv preprint arXiv:2506.06941*, 2025.
- Sarin Sok and Kimkong Heng. ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, 3(1):110–121, 2023.
- Jinrui Tian and Ronghua Zhang. Learners’ AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. *Acta Psychologica*, 260:105725, 2025.
- Vi Loi Truong, Lenh Phan Cong Pham, Nhu Thi Huynh Le, and Luan Trong Nguyen. Factors affecting students’ choices about using ai to enhance individualized education. In *Proceedings of the 2024 7th International Conference on Educational Technology Management*, page 206–212, 2025.
- UNESCO. World declaration on higher education for the twenty-first century: Vision and action, 1998.
- Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, et al. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972): 47–60, 2023.
- Ping Xiao, Yuanyuan Chen, and Weining Bao. Waiting, banning, and embracing: An empirical analysis of adapting policies for generative ai in higher education. *arXiv preprint arXiv:2305.18617*, 2023.
- Rongwu Xu, Brian Lin, Shujian Yang, Tianqi Zhang, Weiyan Shi, Tianwei Zhang, Zhixuan Fang, Wei Xu, and Han Qiu. The Earth is flat because...: Investigating LLMs’ belief towards misinformation via persuasive conversation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, pages 16259–16303, 2024.
- Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1):90–112, 2024.
- Chunpeng Zhai, Santoso Wibowo, and Lily D. Li. The effects of over-reliance on AI dialogue systems on students’ cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1):28, 2024.
- Lexin Zhou, Wout Schellaert, Fernando Martínez-Plumed, Yael Moros-Daval, Cèsar Ferri, and José Hernández-Orallo. Larger and more instructable language models become less reliable. *Nature*, 634(8032):61–68, 2024.

Appendix

A | AI Guidelines for AI Research

We provide detailed guidelines for the eight stages of the research lifecycle, specifically focusing on the discipline of Artificial Intelligence (AI). Our categorization of research stages is informed by the comprehensive framework presented by [Luo et al. \(2025\)](#), which delineates the entire scientific research process into distinct phases including Scientific Hypothesis Discovery, Experiment Planning & Implementation, Paper Writing, and Peer Reviewing. Each task is rated according to our risk-graded framework.

Stage 1: Literature Review

Full Use (★★★★★)

- **Literature Management:** Researchers should fully leverage AI to organize bibliographic entries, generate structured indices based on custom taxonomy, and manage personal reading notes and citations to improve efficiency.

Encouraged Use (★★★★☆)

- **Literature Search:** Under clear research questions and search strategies, AI may be used to execute retrieval tasks to improve recall. The search rules and target directions must be defined by the researcher.
- **Literature Filtering:** AI can filter and classify search results based on user-defined keywords and logic. However, the screening results must be manually reviewed and confirmed by the researcher.

Moderate Use (★★★☆☆)

- **Translation Assistance:** AI may be used to translate foreign literature to lower language barriers. However, key information, terminology, and argumentation details must be verified against the original text.
- **Conceptual Understanding:** When consulting AI for explanations of complex concepts, researchers should treat the output as provisional and scrutinize it against authoritative sources like primary papers and standard textbooks to build a reliable knowledge base.

Cautious Use (★★☆☆☆)

- **Paper Summarization:** AI may generate structured abstracts for single papers to quickly grasp the main points. This must not replace deep reading and full analysis of the text.
- **Paper Comparison:** AI may assist in comparing multiple papers to identify commonalities and differences. The final conclusions must be independently judged by the researcher based on the original texts.

Prohibited Use (☆☆☆☆☆)

- **Deep Reading:** It is prohibited to rely solely on AI summaries for core literature that is critical to the research direction or foundation. Researchers must personally read the full text to understand the argumentation logic, experimental design, and authorial intent.

Stage 2: Research Ideation

Encouraged Use (★★★★☆)

- **Novelty Check:** After forming an initial idea, AI should be used to search for similar existing works to understand the status quo. The final judgment of novelty must be made by the researcher based on systematic reading.
- **Limitation Analysis:** AI may analyze the limitations of existing literature to provide references for refining the research angle.

Moderate Use (★★★☆☆)

- **Proposal Refinement:** Once a clear goal and basic plan are formed, AI can offer structural suggestions to optimize the plan's feasibility and completeness.

Cautious Use (★☆☆☆☆)

- **Research Formulation:** Use caution when using AI to identify unsolved scientific questions. The final condensation and expression of the research question must rely on the researcher's own critical thinking.
- **Research Design:** Use caution when using AI to generate research frameworks or implementation details. The core ideas, innovation points, and logical structure must be led and confirmed by the researcher.
- **Method Selection:** Caution is advised when referencing AI-suggested technical paths. Decisions must be based on the researcher's own judgment of the research goals.

Restricted Use (★☆☆☆☆)

- **Research Assessment:** Avoid using AI to judge the novelty or potential value of a research idea. AI's judgment should not be the sole basis for initiating or abandoning a project; if used, its reasoning must be critically analyzed.

Stage 3: Theoretical Analysis

Encouraged Use (★★★★☆)

- **Theory Review:** AI may be used to survey related theoretical works and model assumptions to expand horizons. The selection and trade-off of theories must be done independently by the researcher.

Moderate Use (★★★☆☆)

- **Proof Assistance:** Once the modeling framework is clear, AI can suggest possible proof directions or lemmas. The researcher must rigorously derive and verify every step.
- **Proof Check:** AI may be used to check proofs for logical gaps or potential counter-examples. The validity of the proof must be confirmed through strict human deduction.

Cautious Use (★☆☆☆☆)

- **Problem Formalization:** Use caution when using AI to mathematize or formalize a research problem. The definition's accuracy and applicability must be confirmed by the researcher based on the problem's essence.
- **Hypothesis Generation:** Use caution when referencing AI-suggested core assumptions. The rationality and necessity of assumptions must be independently evaluated by the researcher.

Prohibited Use (☆☆☆☆☆)

- **Unverified Derivation:** It is prohibited to rely entirely on AI to provide modeling schemes or proofs without personal review, comprehension, and reproduction.
- **Proof Fabrication:** It is prohibited to use AI to generate misleading proofs that mask logical errors to evade academic scrutiny.

Stage 4: Experimentation

Full Use (★★★★★)

- **Boilerplate Code:** For repetitive or template-based code not involving core innovation (e.g., config, visualization, wrappers), AI should be leveraged to reduce labor.

Encouraged Use (★★★★☆)

- **Data Preprocessing:** AI can generate data cleaning and loading logic. The researcher must check the logic's applicability.

- **Documentation:** AI may assist in generating READMEs and install guides to improve reproducibility. The content must be reviewed by the researcher.

Moderate Use (★★★★☆)

- **Experimental Design:** AI may assist in drafting the overall design, control groups, and metrics. The final experimental plan must be decided by the researcher based on scientific judgment.

Cautious Use (★★★☆☆)

- **Data Analysis:** Use caution when asking AI for statistical methods or analysis flows. The choice of method and interpretation of significance must not be delegated to AI.
- **Implementation Assistance:** For modules involving core innovation, AI use is limited to syntax/engineering aid. The researcher must fully understand the core implementation logic.
- **Pipeline Design:** Use caution when using AI to sort out the main experiment execution order. Key procedural decisions must be made by the researcher.

Prohibited Use (☆☆☆☆☆)

- **Data Fabrication:** Strictly prohibited to use AI to falsify, manipulate, or fabricate raw data to match a hypothesis. Any form of result manipulation is severe misconduct.

Stage 5: Scholarly Writing

Full Use (★★★★★)

- **Polishing & Formatting:** AI should be used for grammar checks, symbol consistency, formatting, and language polishing. The researcher must review all changes.

Moderate Use (★★★★☆)

- **Structural Outlining:** AI may assist in generating paper outlines or chapter structures to optimize logic flow.

Cautious Use (★★★☆☆)

- **Related Work Writing:** Based on organized notes, AI may carefully assist in drafting synthetical text (e.g., “Related Work”). Such content must be verified against original sources and rewritten to ensure accuracy.

Restricted Use (★★☆☆☆)

- **Method Writing:** Avoid using AI to generate the first draft of sections involving original methods or models (e.g., “Methodology”). AI output here should serve only as a reference for expression, not the final text.

Prohibited Use (☆☆☆☆☆)

- **Core Contribution Writing:** Strictly prohibited to rely on AI to generate core research viewpoints, innovation arguments, or key reasoning processes.
- **Reference Generation:** Strictly prohibited to use AI to generate bibliography lists or cite unverified references.

Stage 6: Peer Review

Moderate Use (★★★★☆)

- **Reading Assistant:** Reviewers may use AI to summarize papers or check formatting logic to improve efficiency. The final quality assessment must be done independently.

Prohibited Use (☆☆☆☆☆)

- **Review Generation:** It is prohibited to let AI directly summarize issues and generate the review report.

Stage 7: Rebuttal

Encouraged Use (★★★★☆)

- **Rebuttal Organization:** AI may be used to classify and organize comments from multiple reviewers to identify key issues.

Prohibited Use (☆☆☆☆☆)

- **Rebuttal Generation:** It is prohibited to use AI to generate the core response content, including scientific arguments, additional analysis, or defense of the innovation.

Stage 8: Dissemination

Encouraged Use (★★★★☆)

- **Layout Generation:** Researchers are encouraged to use AI tools to quickly generate consistent slide layouts, color schemes, and typographical arrangements based on text drafts.
- **Script Drafting:** It is appropriate to use AI to generate initial speech drafts or key talking points based on the paper's abstract. This serves as an efficiency tool for preparing oral delivery.
- **Mock Q&A:** We encourage using AI agents to simulate audience questions. By feeding the paper's text to an LLM, researchers can engage in multi-turn dialogue to train their ability to respond to critical inquiries.
- **Figure/Poster Design:** AI may be used to design scientific charts, visualize results, and layout conference posters. This includes generating visualization code (e.g., Python scripts) or enhancing aesthetics, provided the underlying data is never altered.

B | AI Guidelines for Teaching and Learning

This section provides detailed guidelines for the application of AI in teaching and learning activities within higher education. The structured stages for teaching and learning reflect the pervasive integration of AI technologies across the educational lifecycle, drawing inspiration from the framework proposed by Kasneci et al. (2023). This framework highlights how AI, particularly large language models, has permeated teachers' pedagogical design (Content Creation & Planning), in-class support (Tutoring), and academic assessment (Assessment & Feedback) throughout the entire educational lifecycle, while also reshaping students' personalized learning paths. Each task is rated according to our risk-graded framework.

1. For Instructors

1.1 Course Preparation

Moderate Use (★★★★☆)

- **Student Needs Analysis:** AI tools may be used to survey students' prior knowledge to assist instructors in understanding their knowledge structure and skill levels. The analysis results should be reviewed and validated by the instructor or teaching assistant as a reference for optimizing teaching plans.
- **Instructional Design:** Instructors should first independently complete the overall design of the course curriculum, clarifying teaching objectives, methods, learning activities, and assessment strategies. Based on this, AI-assisted suggestions for instructional design can be referenced to broaden ideas, refine structure, or complete details.
- **Teaching Material Preparation:** AI may assist in preparing teaching materials, including but not limited to slide decks (PPT), case studies, and practice questions, in accordance with the instructional design. Instructors must rigorously review and correct the accuracy, academic rigor, timeliness, and potential biases of the generated content.

Cautious Use (★★★☆☆)

- **Syllabus Formulation:** AI may assist in generating reference frameworks or structural suggestions for the course syllabus, but the instructor is responsible for defining the scope of course content, scheduling, and pacing.

1.2 Course Delivery

Encouraged Use (★★★★☆)

- **Classroom Recording:** AI tools may be used to record and organize classroom teaching processes, forming digital teaching archives for instructor reflection and student review. Instructors must ensure that the storage, use, and sharing of such data comply with information security and privacy protection requirements.

Moderate Use (★★★★☆)

- **Classroom Participation Analysis:** AI may assist in analyzing classroom interaction (e.g., participation assessment based on voice or video data) to help instructors understand teaching effectiveness and optimize the teaching process. Such applications must prioritize student privacy and data security, and the analysis results should only be used for teaching reference.

Prohibited Use (☆☆☆☆☆)

- **Substitution of Live Instruction:** It is prohibited to use AI-generated virtual audio/video or automated teaching systems to replace systematic instruction by the instructor. The instructor's live guidance, real-time interaction, value transmission, and adaptability in the classroom constitute the core elements of teaching activities and cannot be replaced by AI.

1.3 Post-Class Interaction

Encouraged Use (★★★★☆)

- **AI Teaching Assistant:** AI may be configured based on course-related materials (class recordings, syllabus, teaching slides, Q&A records, etc.) to provide 24/7 assistance to students for post-class practice, review, and answering conceptual or factual questions, and to guide students in Socratic thinking and learning.

Prohibited Use (☆☆☆☆)

- **Substitution of Instructor Office Hours:** AI must not completely replace scheduled instructor office hours to ensure necessary face-to-face communication, intellectual exchange, and academic guidance between instructors and students.

1.4 Course Assessment

Encouraged Use (★★★★☆)

- **Objective Question Grading:** AI may assist in grading objective questions such as multiple-choice, true/false, fill-in-the-blank, and short-answer questions, and in tallying scores. However, instructors must ensure the accuracy of the grading results.

Moderate Use (★★★☆☆)

- **Subjective Question Format Check:** When grading subjective assignments such as essays or reports, AI may assist in checking technical aspects like formatting, language expression, and spelling; however, the evaluation of argumentation logic, academic depth, originality, and research assessment must be completed independently by the instructor.

Cautious Use (★★☆☆☆)

- **Assessment Item Generation:** AI may provide preliminary suggestions for test or assignment questions, but instructors must comprehensively control their academic rigor, difficulty level, and alignment with teaching objectives, and make necessary adjustments.

Prohibited Use (☆☆☆☆)

- **Academic Evaluation:** It is prohibited to solely rely on AI for question generation, grading, and final score determination, or to make critical academic evaluations, capability judgments, and academic development guidance based solely on AI output.

2. For Students

2.1 Pre-Class Preparation

Encouraged Use (★★★★☆)

- **Background Knowledge Acquisition:** AI may be used for information retrieval and organization to quickly understand course-related background knowledge, professional terminology, disciplinary development trends, and cutting-edge advancements, as well as the latest developments and applications of the course topic in academia or industry.
- **Knowledge Framework Structuring:** AI may generate a preliminary course framework, including a list of key knowledge points and suggested learning paths, to assist in building an initial understanding of the course's overall structure and learning priorities.

Moderate Use (★★★☆☆)

- **Pre-Class Reading:** During preview, AI may generate summaries and key points for longer reading materials, but students should subsequently read the original text in its entirety to ensure a comprehensive understanding of the content.

2.2 In-Class Learning

Encouraged Use (★★★★☆)

- **Class Content Comprehension:** During in-class learning, for difficult-to-understand knowledge points or formula derivations, AI may provide instant explanations and analyses to deepen understanding and improve in-class learning efficiency.

- **Class Notes Organization:** AI may assist in generating class notes, course summaries, and knowledge syntheses for organizing learning content. However, students should supplement, revise, and re-process the generated content based on their own understanding, avoiding mechanical acceptance.
- **Classroom Recording:** AI may organize classroom content into reviewable learning materials for post-class review and recall.

Cautious Use (★★☆☆☆)

- **Class Discussion and Presentation:** In segments requiring individual or group thought processes, such as class discussions and group presentations, students should cautiously use AI to generate ideas or answers, to avoid diminishing independent thinking and expressive abilities.

Prohibited Use (☆☆☆☆☆)

- **In-Class Quizzes:** It is prohibited to use AI to directly obtain answers in assessment-oriented segments such as in-class quizzes, exercises, or real-time response systems (e.g., Yu Ketang).

2.3 Post-Class Learning

Full Use (★★★★★)

- **Text Verification and Refinement:** After completing the initial draft of assignments or reports, AI should be used for basic verification and refinement, including spell-check, grammar, and clarity of expression. Students must independently review and finally confirm the modifications suggested by AI.
- **Auxiliary Code Writing:** AI may assist in writing auxiliary code (e.g., data loading, file I/O, environment configuration), provided that students understand the code logic and execution mechanisms.

Encouraged Use (★★★★☆)

- **Daily Q&A:** In post-class learning, for concepts, formula derivations, or problem-solving methods that are not understood, AI may be used for questioning and discussion to obtain multi-perspective explanations and inspirations.
- **Knowledge Consolidation:** AI may organize course slides and key content, generating systematic knowledge summaries, review checklists, and example problems to assist in consolidating understanding and application skills.
- **Exercise Review:** After completing assignments or exercises, AI may be used to compare and analyze problem-solving approaches and methods, deepening the understanding of the solution process.
- **Code Inspection and Debugging:** AI may assist in checking for code errors and understanding debugging strategies to enhance problem identification and debugging skills.
- **Assignment Check:** After completing an assignment, AI may be used to check the results, but students must understand the complete solution process themselves.

Moderate Use (★★★☆☆)

- **Problem-Solving Hints:** When encountering difficult problems, students may ask AI for hints on solution directions, relevant theorems, or method suggestions, but the complete derivation or implementation process should be completed independently by the student.

Prohibited Use (☆☆☆☆☆)

- **AI Ghostwriting of Assignments:** It is prohibited to submit assignment requirements directly to AI to generate complete answers and submit them as personal work, including but not limited to mathematical proofs, derivation processes, calculation steps, or core code implementations.

2.4 Course Assessment

2.4.1 Major Assignments/Projects

Encouraged Use (★★★★☆)

- **Literature Survey:** AI may assist in searching for relevant academic papers, technical documents, and cutting-edge developments to provide references for project research.
- **Feasibility Analysis:** AI may analyze the technical difficulties, resource requirements, and potential risks of a project idea as an auxiliary tool for proposal argumentation.
- **Project Management:** AI may assist in task breakdown, timeline planning, and progress tracking suggestions to improve project execution efficiency.

Moderate Use (★★★☆☆)

- **Technical Solution Exploration:** When encountering technical difficulties, students may seek AI for problem-solving ideas, algorithm suggestions, or tool recommendations.

Cautious Use (★★☆☆☆)

- **Report Writing:** AI may generate a report framework, providing suggestions for report structure and chapter highlights for reference.
- **Topic Selection:** AI may offer various possibilities for project ideas, research directions, or technical solutions, but students must evaluate the project's originality, feasibility, and learning value themselves.

Prohibited Use (☆☆☆☆☆)

- **Core Content Implementation:** It is prohibited to allow AI to complete the core algorithms, key logic, or innovative aspects of a project, as this part is central and critical to achieving course objectives and demonstrating learning outcomes.

2.4.2 Examinations and Quizzes

Prohibited Use (☆☆☆☆☆)

- **Examinations and Quizzes:** It is prohibited to use AI tools to obtain answers or assist in answering in any form of examination, quiz, or assessment.
- **Assessments:** It is prohibited to use AI in assessment tasks that require independent completion, including independent programming assessments and live oral defenses.

C | Risk-Graded AI Compliance Checklist

This checklist is designed to operationalize the risk-graded guidelines proposed in this paper. It ensures that the efficiency gains from AI tools do not come at the expense of scientific integrity or cognitive responsibility. Authors must answer the following questions regarding their usage of AI across the research lifecycle (Stages 1–5).

For each question pair in the checklist:

- Authors should answer [Yes], [No], or [N/A].
- [N/A] should be selected when a question does not apply to the paper. For the first question (e.g., “Did you use AI...?”) in each pair, choose [N/A] if you did not perform the relevant task or activity at all. For the second question (e.g., “If yes, did you...?”), choose [N/A] if AI was not used for that task or if the task itself was not performed.
- While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so we suggest authors using their best judgment and write a justification to elaborate. It is recommended to provide a short (1–2 sentence) justification right after the answer to the first question (even for [N/A]).

To provide a concrete example, we have filled in the checklist below based on the actual AI usage in this paper.

1. If you conducted a literature review...

- Did you use AI for literature search or filtering? [Yes] **Justification:** AI was used to assist with keyword expansion and initial paper retrieval.
If yes, did you manually review the results for relevance and completeness? [Yes]
- Did you use AI to translate text or explain complex concepts? [Yes] **Justification:** AI was used to help interpret technical descriptions and non-native-language materials.
If yes, did you verify terminology and explanations against the original sources? [Yes]
- Did you use AI to read or summarize the core literature essential to your research foundation? [Yes] **Justification:** AI was used for preliminary summarization to aid navigation of the literature.
If yes, did you personally confirm the key claims and conclusions from the original papers? [Yes]
- Did you use AI to compare multiple papers? [Yes] **Justification:** AI was used to highlight potential overlaps and contrasts across related work.
If yes, did you verify the similarities and differences using the original texts? [Yes]

2. When you proposed research ideas...

- Did you use AI to assess the value or significance of your research idea? [No] **Justification:** The research motivation and significance were determined entirely by the authors.
If yes, did you make the final judgment based on your own independent evaluation? [N/A]
- Did you use AI to check the novelty or limitations of existing work? [No] **Justification:** Novelty and limitations were assessed through manual review of the literature.
If yes, did you manually verify the cited literature against the original sources? [N/A]
- Did AI suggest technical approaches or frameworks for your research? [No] **Justification:** Our research framework was determined entirely by the authors.
If yes, did you assess feasibility and alignment with your research goals? [N/A]
- Did you use AI to help formulate the research question or hypothesis? [No] **Justification:** The research questions were formulated independently by the authors.
If yes, did you independently finalize the research question and hypothesis? [N/A]
- Did you use AI to review the research plan for logical consistency? [No] **Justification:** Logical consistency was ensured through internal discussion and human review.
If yes, did you personally check the plan to ensure all reasoning was sound? [N/A]

3. If you performed theoretical analysis...

- (a) Did you use AI to draft theoretical assumptions or formal problem definitions? [N/A] **Justification:** This paper does not involve formal theoretical assumptions or definitions. If yes, did you independently verify their rationality and necessity? [N/A]
- (b) Did AI suggest proof strategies, lemmas, or derivation steps? [N/A] **Justification:** No formal proofs or derivations were part of this work. If yes, did you manually derive and verify every step to ensure correctness? [N/A]
- (c) Did you use AI to check or review proofs? [N/A] **Justification:** The paper does not contain formal proofs. If yes, did you perform a full human review to confirm validity and identify gaps? [N/A]
- (d) Did you use AI to help understand or reproduce theoretical results? [N/A] **Justification:** There were no theoretical results requiring reproduction. If yes, did you personally verify all results and derivations? [N/A]

4. If you conducted experiments...

- (a) Did you use AI in generating, analyzing, or interpreting experimental data or empirical results? [N/A] **Justification:** This work does not include experimental or empirical data. If yes, did you ensure all data and results were authentic? [N/A]
- (b) Did AI provide suggestions for experimental design or evaluation metrics? [N/A] **Justification:** No experimental design was involved in this study. If yes, did you make the final design decisions based on your own scientific judgment? [N/A]
- (c) Did AI assist in writing code related to core experimental contributions? [N/A] **Justification:** The paper does not include experimental code. If yes, did you fully review and understand the implementation logic? [N/A]
- (d) Did you use AI for data preprocessing or statistical analysis? [N/A] **Justification:** No data preprocessing or statistical analysis was conducted. If yes, did you manually verify the correctness and applicability of the methods used? [N/A]
- (e) Did you use AI to generate auxiliary materials (e.g., documentation or code scaffolding)? [N/A] **Justification:** No auxiliary materials of this type were generated. If yes, did you thoroughly review their accuracy and functionality? [N/A]

5. When you wrote the paper...

- (a) Did you use AI to help determine the core viewpoints, arguments, or reasoning of the paper? [No] **Justification:** The central arguments and reasoning were developed solely by the authors. If yes, did you personally verify and establish the main arguments and conclusions? [N/A]
- (b) Did AI assist in drafting sections involving original methods or models? [Yes] **Justification:** AI was used to refine wording and structure after human-authored drafts were completed. If yes, were the initial drafts produced by humans and AI used only for refinement? [Yes]
- (c) Did AI assist in synthesizing related work? [Yes] **Justification:** AI was used to help organize and summarize existing literature during writing. If yes, did you verify the content against the original sources to ensure accuracy? [Yes]
- (d) Did AI assist in language polishing or formatting? [Yes] **Justification:** AI was used for grammar checking and stylistic polishing. If yes, did you review the final text to ensure semantic fidelity and scientific precision? [Yes]
- (e) Did you use AI to assist with generating citations in the bibliography? [No] **Justification:** All citations were selected and verified manually by the authors. If yes, did you manually verify the existence and accuracy of every citation? [N/A]

D | Discussions

Detailed comparison to recent conference guidelines. We found that AI use guidelines in major conferences have become more detailed and considered the corresponding risks, but they still leave an important gap that our framework aims to fill. A recent example is the NeurIPS 2026 Position Paper Track policy⁷, which explicitly recognizes both the productivity benefits and integrity risks of AI use. The policy states:

While we recognise the productivity gains that can be realised through judicious use of AI in research, due to the risk to the integrity of individual projects and of the review system as a whole, the position paper track is establishing the following explicit guardrails on AI use in preparing and reviewing submissions.

- While AI tools may be used in the research that leads to the final paper, the final paper must itself be substantially written by human authors, meaning that AI is used only for copy-editing or similar peripheral changes to the main text.
- At submission time, authors will be required to state how AI tools were used in the preparation of the paper, if at all, and to attest that they have not used AI in ways contrary to the above rule.
- Because papers submitted to the position paper track are confidential, reviewers will be required to commit to not using AI tools to write their reviews.
- Reviewers and authors found to have contravened their commitments not to use AI may be subject to desk-rejection of any work submitted to the position paper track.
- Note that the Position Paper Track’s LLM policy differs from the Main Program’s LLM policy. Authors are responsible for understanding policy pertaining to the specific track they are submitting to, and abiding by it.

This policy is notably more detailed than earlier broad disclosure-based guidance. It distinguishes between AI use in the research process and AI use in final paper writing, requires disclosure and attestation at submission time, prohibits reviewers from using AI tools to write confidential reviews, and explicitly notes that LLM policies may differ across conference tracks. These developments show that conference policies are becoming more context-aware.

However, the policy also illustrates the remaining gap between current guidance and our proposal. It provides venue- and track-specific guidelines for preparing and reviewing submissions, but it does not offer a general task-level taxonomy for higher education activities such as literature review, problem formalization, experimental design, proof checking. Our framework is therefore complementary: it translates broad principles such as disclosure, verification, confidentiality, and human responsibility into a risk-graded structure that distinguishes concrete tasks and stages of academic work. Thus, although AI use guidelines have become more detailed this year, a gap remains between current guidance and our proposed risk-graded framework. Our contribution is therefore not simply to advocate that different venues set different rules, but to provide a more fine-grained and systematic taxonomy that can support such rule-setting in a clearer and more actionable way.

The generalizability of the framework. Although our detailed guidelines use AI research and teaching as an illustrative case, the framework is not specific to AI as a discipline. Its generalizability comes from its grounding in the shared mission of higher education: cultivating students’ capabilities, judgment, originality, and academic responsibility. Across disciplines, AI-related risks such as factual errors, over-reliance, reduced independent thinking, privacy violations, and academic misconduct can arise whenever AI tools are used in learning, research, writing, or assessment. What varies across disciplines is the concrete form of the activity, not the underlying governance logic.

Accordingly, the AI research case study should be understood as one instantiation of the framework rather than the boundary of its applicability. The specific mapping of tasks to risk levels may not transfer directly to fields such as medicine, law, humanities, or engineering, but the procedure does: institutions can identify discipline-specific tasks, evaluate the relevant risks, and assign corresponding levels of AI use. In this sense, what generalizes is not the AI-specific checklist itself, but the risk-graded logic that structures decision-making. Appendix C further illustrates this broader orientation by including teaching and learning guidelines, showing that the framework is intended for multiple higher-education activities rather than a single research workflow.

⁷<https://neurips.cc/Conferences/2026/CallForPositionPapers>

Malicious misuse. Our framework is not intended as an auditing or detection system. A checklist cannot prevent malicious misuse, since intentional violations may simply go undisclosed. Rather, its role is similar to conference disclosure requirements, ethics statements, or data availability statements: it clarifies expectations, standardizes reporting, and supports human oversight. The goal is therefore not to determine whether AI was used at all, but to guide how AI is used responsibly, transparently, and within appropriate bounds.

This limitation follows from the nature of our proposal. Educational institutions routinely articulate norms for academic integrity and research ethics even though full compliance cannot be guaranteed. Similarly, our checklist and risk-graded framework do not replace auditing, enforcement, or institutional sanctions where needed. Instead, they reduce misuse caused by ambiguity, encourage reflection before AI use, and provide a structured mechanism for communicating responsible AI involvement in academic work.