

The Structural Risks of AI in Higher Education: From Integrity Discourse to Epistemic Reconfiguration

Zhuyu Wang *

Sichuan Normal University, Chengdu, China

* Corresponding Author Email: 970106029@qq.com

Abstract. Debates about artificial intelligence in higher education often focus on individual misconduct and academic integrity. This paper examines risks that arise when AI becomes embedded in the infrastructures, routines, and governance of universities. Through a conceptual analysis informed by existing literature, it considers surveillance and data extraction, algorithmic bias, epistemic convergence, cognitive deskilling, relational displacement, and institutional dependence on technological systems. It argues that these risks are reinforced by commercial acceleration, delayed governance, and tensions between educational values and platform priorities. The discussion develops responses centred on institutional accountability, assessment redesign, critical AI literacy, and the protection of educational agency. It also considers the practical constraints of resources, expertise, and institutional capacity. The paper concludes that responsible AI integration requires attention to how technologies reshape knowledge, academic work, and educational relationships, alongside safeguards for individual users.

Keywords: Artificial intelligence; higher education; structural risk; educational agency; algorithmic governance; critical AI literacy.

1. Introduction

The emergence of generative artificial intelligence in higher education has provoked a response that, in much institutional and public discourse, remains centred on academic dishonesty. As universities increasingly integrate AI into assessment, teaching, and student support systems, concerns about academic integrity have become the dominant lens through which technology is understood. This essay focuses specifically on higher education, where the adoption of generative AI intersects directly with production, evaluation, and certification of knowledge. However, framing AI primarily as a problem of cheating is analytically insufficient. It treats AI as an external threat to an otherwise coherent educational system, positioning the problem as individual moral failure susceptible to technical detection and punitive enforcement. What it obscures is a deeper institutional crisis — one concerning the purposes of assessment, the conditions of knowledge production, and the governance of technologies whose development is driven by commercial rather than educational imperatives.

Scarfe et al. [1] provides perhaps the most striking empirical illustration of this misdirection. In a blind study at a UK university, they injected fully AI-generated submissions into five undergraduate modules and found that 94% went undetected by institutional systems. More significantly, AI submissions achieved grades on average half a classification boundary higher than those of real students, with an 83.4% probability of outperforming a random selection of human submissions. This finding deserves more than procedural attention. If AI can consistently outperform students under standard assessment conditions, the question shifts from "how to catch cheating" to "what existing assessments are actually measuring"— and, by extension, what cognitive labour universities are actually asking students to perform. Garzón et al. [2], reviewing 155 empirical studies published between 2015 and 2025, find that the field remains dominated by short-term implementation studies, with limited systematic attention to long-term pedagogical or ethical consequences. The reactive dynamic observable in scholarship mirrors the reactive dynamic in institutional policy.

This essay develops a structural argument. Drawing on critical technology theory [3] and critiques of surveillance capitalism [4], it argues that generative AI should be understood not merely as a tool

deployed within education, but as part of a broader optimisation-driven infrastructure that reorganizes the conditions under which knowledge, authority, and assessment are produced. The following sections establish a working definition of "structural" as a critical concept, analyse structural risk across infrastructural, pedagogical, and governance dimensions, and examine their interconnected consequences. The essay then evaluates existing governance responses and proposes mitigation strategies grounded in assessment redesign.

2. Defining “Structural”: A Conceptual Framework

Before analysing specific risks, it is necessary to clarify what it means to call them “structural.” The term distinguishes a particular class of problems from two more familiar framings: technical failures, which locate the source of harm in malfunctions intrinsic to a given system, and ethical-individualist failures, which locate it in the choices of individual actors. Structural problems, by contrast, are produced by the arrangements within which technologies are developed, deployed, and governed through institutional incentives, organisational roles, and distributional effects that shape how a technology functions regardless of the intentions of any individual user.

Applied to AI in education, three structural tensions are of primary concern. The first is commercial acceleration: the leading AI platforms are developed by firms whose primary obligations are to investors rather than to educational communities, and whose expansion into educational markets prioritises scale and speed over pedagogical fitness. Feenberg [3] describes this as technological rationalization, a process through which institutional practices gradually adapt to the internal logic of external systems, marginalising values that are incompatible with optimisation-based design. Pedagogical values such as interpretive ambiguity, intellectual struggle, and careful slow reasoning are not incidentally neglected by commercial AI; they are structurally incompatible with platforms designed for friction-free output and user retention.

The second tension is governance lag. Regulatory and institutional responses have consistently operated reactively. Garzón et al. [2] document how current research remains dominated by short-term intervention studies, without robust evidence concerning long-term pedagogical or ethical consequences — a reflection, at the level of scholarship, of the same dynamic found in policy. UNESCO [5] likewise acknowledges that the majority of educational institutions have yet to develop formal AI governance frameworks, confirming that adoption is outpacing oversight. AI systems evolve continuously through cloud-based deployment, while educational governance operates through slow bureaucratic cycles of consultation and approval. The asymmetry is not temporary; it is structural.

The third tension is value incompatibility. The optimisation targets of commercial AI such as engagement, efficiency, data extraction, and scalability are frequently incompatible with educational values. An educational system oriented toward critical inquiry, relational learning, and equitable development does not naturally emerge from infrastructure designed to maximise platform growth. This framework shifts the evaluative focus. Rather than asking whether AI tools are being used responsibly by individuals, it asks whether the conditions under which those tools are adopted are compatible with educational ends. The three sections that follow examine this question at the infrastructural, pedagogical, and institutional levels respectively.

3. Structural Risks as Infrastructural Effects

3.1. Surveillance, Data Extraction, and Opacity

AI-driven educational systems increasingly transform classrooms into environments of passive surveillance. Student writing, feedback interactions, and behavioural traces are collected at scale and transformed into training data for model refinement. In a systematic review of 75 empirical papers, Zhu et al. [6] identified privacy invasion as the most frequently documented risk, appearing in 46 papers, closely followed by data leakage in 32 papers. Privacy invasion and data leakage are

conditions built into the commercial model of AI platforms. Large-scale collection of personal data — including language use and interaction patterns — occurs often without meaningful informed consent precisely because it is commercially necessary. This dynamic exemplifies what Zuboff [4] terms surveillance capitalism: the systematic extraction of behavioural data as a resource for commercial prediction and control.

Compounding this is the opacity of black-box architectures, documented in 29 of 75 papers by Zhu et al. [6]. Because the deep-learning mechanisms of large language models remain inaccessible to inspection, educators are placed in a structurally contradictory position: they remain legally and ethically responsible for grading, feedback, and student evaluation decisions that may be substantially mediated by automated systems they can neither audit nor interpret. This is not a contingent problem awaiting a technical fix. It reflects the design logic of commercial platforms, which are not structured to provide the kind of transparency that educational accountability requires.

3.2. Algorithmic Bias

Algorithmic bias represents the single most prevalent risk in recent literature, identified in 53 of 75 papers by Zhu et al. [6]. Bias is not a computational error but a structural outcome of training on social and linguistically stratified data. The biases encoded in training corpora reflecting existing racial, gender, and socioeconomic inequalities are not merely reproduced when AI systems make consequential decisions about admissions, grading, or learning provision; in some cases, they are amplified through feedback loops, as systems trained on historical patterns reinforce the conditions that generated those patterns. This is a structural feature of optimisation-based design, not a governance failure that could be resolved through better monitoring alone.

The European Parliament's AI Act [7] acknowledges these systemic hazards, classifying AI used in education and vocational training as “high-risk” and explicitly banning emotion-inference systems in classrooms. This regulatory concession that the risks are systemic in nature is significant. However, the Act's full compliance obligations do not apply until 2027, and its sector-agnostic legal architecture fails to address the specific epistemic harm that bias inflicts on marginalized student populations. The regulatory intent is real; the operational gap between intent and implementation is itself a structural feature of the current governance landscape.

4. Pedagogical Reconfiguration: Knowledge, Labour, and Relation

The structural embedding of optimisation logics does not merely produce discrete risks; it reconfigures the conditions under which teaching and learning occur. Three interconnected transformations are worth examining in detail, because each one strikes at a different foundational assumption of higher education.

4.1. Epistemic Convergence and the Narrowing of Inquiry

It is worth acknowledging that generative AI can broaden access to knowledge — reducing barriers for students who previously lacked specialist texts or writing support, and any account of epistemic convergence that ignores this risks overstating the case.

The concern, however, is not that convergence is inevitable but that existing platform architectures create strong structural incentives in that direction. As AI text generators and algorithmic recommendation engines increasingly mediate access to information, students are guided toward pre-structured, automated pathways of interpretation. The mechanism is subtle. It does not resemble censorship in any conventional sense. Rather, it works through the gradual normalisation of high-probability cognitive routes: the statistical tendency of large language models to produce outputs that reflect the most common patterns in their training data. Knowledge, as a result, is encountered not as an open-ended and frequently uncomfortable inquiry but as a set of pre-optimized trajectories.

Li [8] argues that generative AI shifts the focus of curriculum from open human exploration toward closed, standardized knowledge synthesis. The practical implication is a narrowing of intellectual

range: minority perspectives, heterodox arguments, and historically marginalised epistemologies are structurally underrepresented in the outputs of systems trained on dominant corpora. This does not happen through any deliberate act of exclusion; it happens through the mathematics of probability. The question this poses for higher education is whether institutions are willing to actively counteract that tendency — through curriculum design, assessment practice, and critical pedagogy — or whether they will allow the intellectual range of student inquiry to drift toward whatever the algorithm judges most likely.

4.2. The Deskilling of Cognitive Labour

Concerns about the reduction of intellectual labour are not new. Apple and Jungck [9] warned that educational technologies could gradually shift professional work away from judgement and towards the execution of externally designed procedures, a framework that gains renewed urgency in the context of generative AI.

Recent developments have intensified these concerns. Scarfe et al. [1] demonstrated that AI systems can produce sophisticated written arguments independently of human cognition, often achieving higher grades than undergraduate students under conventional assessment conditions. The significance of this finding lies not in any comparison between human and machine performance, but in what it reveals about the structural assumptions embedded within standard assessment practices.

The question is whether those assumptions can still guarantee the cognitive engagement that higher education has traditionally sought to cultivate. Writing is not merely a means of communicating knowledge; it is also a process through which knowledge is developed. Activities such as synthesising evidence, evaluating competing interpretations, and constructing sustained arguments require intellectual effort that contributes directly to learning. When these processes are increasingly delegated to AI systems, students may engage less frequently with the forms of cognitive labour that writing has historically supported.

This does not mean that the use of AI necessarily undermines learning. In some contexts, generative tools may assist brainstorming, provide formative feedback, or help students overcome barriers to participation. Vetter et al. [10], examining a composition student's use of ChatGPT, similarly find that AI can function as a productive brainstorming and drafting partner, yet note that when students delegate structural argumentation to AI, they risk diminishing the cognitive agency that writing is meant to develop — a tension they describe as a core pedagogical ethical concern. Nevertheless, the widespread availability of systems capable of producing coherent academic outputs raises important questions about what assessments are actually measuring. If a substantial proportion of the intellectual work associated with writing can be externalised, then conventional take-home essays may become less reliable indicators of understanding, development, and independent judgement.

From a structural perspective, the challenge lies not in preventing students from using AI, but in ensuring that assessment continues to capture meaningful evidence of learning in an environment where cognitive labour can increasingly be automated.

4.3. Relational Displacement and the Automation of Care

Perhaps the most difficult consequence to measure is the effect of AI on educational relationships. Biesta [11] argues, from a philosophical standpoint, that education is irreducibly an ethical and non-instrumental encounter, one in which students are not simply trained toward specified outcomes but invited into a form of existence that involves genuine responsibility for others. On this view, pedagogical authority cannot be reduced to the efficient delivery of information, and the teacher-student relationship cannot be replaced by a system that dispenses optimized feedback.

Tong and Li [12], examining the philosophical dimensions of AI's impact on teacher roles, make a related point about the embodied nature of this encounter. They argue that the emotional and relational dimensions of teaching — what they describe as the "embodied" quality of care and attention — cannot be replicated by AI systems, whose responses to students are, at bottom,

probabilistic outputs rather than genuinely intentional replies to a specific human being. This is not a romanticization of teaching; it is a claim about what educational relationships actually do. When generative AI is positioned to automate feedback, personal guidance, and even assessment, the teacher-student dyad does not simply become more efficient. Educational interactions are reorganized around platform-mediated logistics, transforming what Tong and Li describe as an ethical human encounter into an instrumental data transaction. More fundamentally, widespread automation may alter the purpose of educational encounters themselves.

5. Governance Misalignment and Structural Lock-in

Attempts to regulate AI in education primarily operate at the level of individual usage, behavioural guidelines, and compliance checklists. Transnational frameworks such as the UNESCO [5] guidelines and regional statutes like the EU AI Act [7] seek to ensure transparency, accountability, and fairness in deployment. These efforts are not without value. But they suffer from a deep structural misalignment: they fail to address the political economy that drives institutional dependence on corporate AI infrastructures in the first place.

The optimistic counterposition deserves acknowledgement: universities retain the capacity to assert control over technology vendors and enforce institutional ethical codes, and some do implement local frameworks to audit the technologies they use.

This argument, however, may underestimate the extent to which institutional choices are constrained by existing technological infrastructures. Selwyn [13, 14] describes this dynamic through the concepts of vendor lock-in and path dependency. Once digital systems become embedded within assessment platforms, virtual learning environments, administrative workflows, and student support services, replacing them often requires substantial financial investment, technical expertise, and organisational change.

For example, universities that integrate commercial AI tools across multiple institutional functions may find themselves increasingly dependent on proprietary ecosystems that are difficult to audit and even more difficult to replace. Zheng [15] further argues that the ethical governance of generative AI is complicated by the opacity of proprietary models, limiting meaningful institutional oversight even where a clear commitment to accountability exists.

The result is a gradual movement toward structural lock-in, where institutions become increasingly occupied with managing the consequences of technological dependence rather than shaping the conditions under which technologies are adopted. This process emerges incrementally through successive integrations, accumulated dependencies, and growing switching costs.

This trajectory is not irreversible, but reversing it requires more than stronger compliance mechanisms. Governance reforms that operate at the level of individual usage policies or ethical guidelines leave the underlying infrastructural dependency intact. Any substantive response will therefore need to address not only how AI tools are used within universities, but how educational institutions acquire, govern, and where possible develop the technological infrastructures on which they depend. Without that reorientation, the dominant direction of travel is one in which regulatory frameworks increasingly manage the terms set by commercial monopolies rather than challenging the conditions that produced those terms.

6. Mitigating Structural Risk: Reclaiming Educational Agency

If the risks analysed above are structural rather than merely behavioural, mitigation cannot be confined to plagiarism detection or usage policies. Effective responses must instead address the institutional conditions under which AI is adopted.

The most immediate site of intervention is assessment redesign. The Higher Education Policy Institute [16] found that 88% of UK undergraduates had used generative AI for assessments by 2025, up from 53% the previous year, a pace that makes detection-first approaches structurally inadequate.

Mitigation requires a shift from product-oriented to process-oriented evaluation: oral defence, iterative drafting, and reflective commentary that keep cognitive labour visible. The trade-off is real: oral components increase marking workload and may disadvantage students with weaker spoken fluency. The goal, however, is not to exclude AI from education, only to ensure that the intellectual work assessment is meant to evaluate cannot be invisibly externalised.

Institutional governance represents a second, deeper layer. Fudan University's Regulations on the Use of AI Tools in Undergraduate Theses [17] illustrates what substantive governance looks like in practice: the regulations distinguish structurally between AI as procedural aid — permitting use in literature retrieval and citation formatting — and AI as substitute for cognitive labour, prohibiting its application to research design, argument construction, and written composition. The document also prohibits assessment panels from using AI to evaluate student work, closing a loop many frameworks leave open. Yet even this framework remains partly detection-oriented, leaving the deeper question of epistemic dependence unresolved. Moreover, its implementation depends on substantial institutional capacity and may therefore be more difficult for less well-resourced universities to replicate consistently.

The third layer is AI literacy reoriented toward critical understanding rather than procedural compliance. HEPI [16] found that students' primary deterrent to AI use was fear of misconduct accusations rather than any critical grasp of AI's limitations, suggesting that current literacy efforts remain behavioural rather than epistemic. The aim is not to reject AI but to preserve the capacity of educational values to shape its adoption, rather than being subordinated to it.

7. Conclusion

The dangers of AI in higher education cannot be adequately understood through the narrow lens of academic misconduct or individual misuse. As this essay has argued, the most significant risks are structural and, crucially, they are mutually reinforcing. Surveillance and data opacity do not operate in isolation. As student writing and interaction patterns are continuously collected to refine commercial models, the outputs of those models increasingly reflect dominant patterns in aggregated data, creating a feedback loop that intensifies epistemic convergence. Students guided by systems trained on high-probability cognitive routes are less likely to encounter the heterodox and marginalised perspectives on which critical inquiry depends. Deskilling therefore becomes a structurally predictable consequence rather than an incidental outcome. Where governance lock-in prevents institutions from auditing or replacing the platforms that generate these effects, this process becomes increasingly difficult to reverse. The problem arises not from any single institutional decision but from the gradual accumulation of technological dependency.

Assessment redesign preserves visibility of cognitive labour but increases marking workloads and may disadvantage students with weaker oral fluency. AI literacy programmes remain difficult to evaluate and risk becoming procedural compliance exercises rather than genuinely critical orientations.

What none of these strategies alone addresses is a deeper institutional problem: the conditions under which universities decide what AI is for. So long as adoption is driven primarily by efficiency pressures and competitive visibility rather than pedagogical deliberation, structural risks will persist regardless of which mitigation mechanisms are in place. The future of education as a human practice depends not only on how AI is regulated, but on whether universities retain the institutional will to ask that question at all.

References

- [1] Scarfe, P., Watcham, K., Clarke, A., & Roesch, E. (2024). A real-world test of artificial intelligence infiltration of a university examinations system: A "Turing Test" case study. *Plos One*, 19(6), Article e0305354. <https://doi.org/10.1371/journal.pone.0305354>

- [2] Garzón, J., Patiño, E., & Marulanda, C. (2025). Systematic Review of Artificial Intelligence in Education: Trends, Benefits, and Challenges. *Multimodal Technologies and Interaction*, 9(8), Article 84. <https://doi.org/10.3390/mti9080084>
- [3] Feenberg, A. (2002). *Transforming technology: A critical theory revisited*. Oxford University Press. <https://doi.org/10.1093/oso/9780195146158.001.0001>
- [4] Zuboff, S. (2018). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.
- [5] UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization. <https://unesdoc.unesco.org>
- [6] Zhu, H., Sun, Y., & Yang, J. (2025). Towards responsible artificial intelligence in education: A systematic review on identifying and mitigating ethical risks. *Humanities and Social Sciences Communications*, 12(1), Article 1111. <https://doi.org/10.1057/s41599-025-05252-6>
- [7] European Parliament. (2024). *Artificial Intelligence Act: Regulation laying down harmonised rules on artificial intelligence*. <https://www.europarl.europa.eu>
- [8] Li, J. (2024). The impact of generative AI on curriculum, teaching materials, and pedagogical methods (Shēngchǎnshì réngōngzhìnéng duì kèchéng jiàocái jiàofǎ de yǐngxiǎng). *Curriculum, Teaching Material and Method (Kèchéng, Jiàocái, Jiàofǎ)*, 44(2), 15–22. <https://doi.org/10.19877/j.cnki.kcjcjf.2024.02.015>
- [9] Apple, M. W., & Jungck, S. (1990). “You Don’t Have to Be a teacher to Teach This Unit.” *Teaching, Technology, and Gender in the Classroom*. *American Educational Research Journal*, 27(2), 227–251. <https://doi.org/10.3102/00028312027002227>
- [10] Vetter, M. A., Lucia, B., Jiang, J., & Othman, M. (2024). Towards a framework for local interrogation of AI ethics: A case study on text generators, academic integrity, and composing with ChatGPT. *Computers and Composition*, 71, Article 102831. <https://doi.org/10.1016/j.compcom.2024.102831>
- [11] Biesta, G. (2017). *The rediscovery of teaching*. Routledge. <https://doi.org/10.4324/9781315617497>
- [12] Tong, J., & Li, M. (2026). Empowering teachers through generative AI: A philosophical examination and future directions (Shēngchǎnshì réngōngzhìnéng fùnéng jiàoshī juésè: Zhéxué shěnsì yǔ wèilái zǒuxiàng). *Theory and Practice of Education (Jiàoyù Lílùn yǔ Shíjiàn)*, 46(5), 22–28.
- [13] Selwyn, N. (2014). *Distrusting educational technology: Critical questions for changing times*. Routledge. <https://doi.org/10.4324/9781315886350>
- [14] Selwyn, N. (2016). *Is technology good for education?* Polity Press. <https://research.monash.edu/en/publications/is-technology-good-for-education>
- [15] Zheng, H. (2025). Research on the ethical risks and trustworthy governance pathways of generative AI (Shēngchǎnshì réngōngzhìnéng de lúnlǐ fēngxiǎn yǔ kěxìn zhìlǐ lùjìng yánjiū). *Science & Technology Progress and Policy (Kējì Jìnbù yǔ Duìcè)*, 42(12), 38–48.
- [16] Higher Education Policy Institute. (2025). *Student Generative AI survey 2025 (HEPI Policy Note 61)*. <https://www.hepi.ac.uk/reports/student-generative-ai-survey-2025/>
- [17] Fudan University. (2024). *Regulations on the use of AI tools in undergraduate theses (trial) (Fùdàn Dàxué Guānyú Zài Běnkē Bìyè Lùnwén (Shèjì) Zhōng Shǐyòng AI Gōngjù de Guǐdìng (Shíxíng))*. Office of Academic Affairs. <http://www.it.fudan.edu.cn/Data/View/5005>