

Cognitive Outsourcing in Digital-Intelligent Education: Hidden Educational Risks and Regulatory Paths

Zhiyue Jin

Sichuan Normal University, Chengdu, Sichuan, China

2029561093@qq.com

Abstract. The deep integration of artificial intelligence in classroom teaching reorganises learners' internal cognition and technologically mediated external cognition through the mechanism of cognitive offloading, fundamentally reshaping how knowledge is accessed and thinking is processed. Yet algorithmic limitations and misaligned learner orientations are driving this offloading toward risks of hollowed classroom interaction, flattened cognitive depth, and distorted feedback. This study systematically classifies these risks to fill a critical gap. Based on a four-dimensional framework of thinking, subject, knowledge, and practice, the risks are categorised as follows: at the thinking dimension, algorithmic opacity and data bias reduce instruction from thinking cultivation to answer retrieval, obstructing higher-order goals through a black-box trap that masks reasoning processes; at the subject dimension, the quasi-subjectivity of intelligent systems encroaches upon the learner's agency, triggering subject alienation and hollowing out authentic classroom interaction; at the knowledge dimension, fragmented algorithmic outputs are mistaken for systematic understanding, leaving comprehension at a superficial level; at the practice dimension, instant system responsiveness dissolves productive ambiguity and interrupts situated practical activities, disrupting human-machine synergy in situated learning. Through analysing these risks, this study proposes mitigation strategies across teaching, learning, and assessment: rebuilding classroom thinking space through generative dialogue between teachers and students; requiring independent learner reasoning before critical scrutiny of algorithmic outputs; and establishing a process-trace-centred evaluation mechanism that values drafts and revision histories to reflect genuine thinking development.

Keywords: Artificial Intelligence, Teaching Application, Cognitive Outsourcing, Human-Machine Collaboration.

1. Introduction

1.1. Policy Requirements and Practical Needs

China has positioned artificial intelligence at the heart of its educational modernisation strategy. On 10 April 2026, China's Ministry of Education and six other central departments issued the "AI+Education" Action Plan (Document No. Jiaokexin [2026] 1), mandating AI curricula from primary to higher education [1]. Significantly, the plan explicitly identifies risks including "cognitive inertia" and "excessive reliance" on AI, calling for preventive measures to ensure healthy development. In the United Kingdom, the Department for Education updated its Generative AI: Product Safety Standards on 19 January 2026 [2]. The standards adopt a progressive disclosure approach to prevent cognitive deskilling and prohibit manipulative design tactics that could undermine learner agency. Both policies frame AI as an educational enabler while implicitly acknowledging the cognitive dangers that this paper systematically examines.

Yet beneath this efficiency narrative lies a quieter redistribution of epistemic labour. Current systematic reviews remain preoccupied with procedural concerns such as data privacy, algorithmic bias, and academic integrity [3, 4], while overlooking how habitual delegation of memory and metacognitive monitoring reshapes the learner's intellectual subjectivity [5]. This oversight matters because, as Chinese scholars have observed, pre-existing cognitive inertia, not AI itself, creates the conditions for AI misuse [6]. This risk emerges from the interplay between learners' cognitive inertia and AI's convenience and accessibility.

1.2. Thesis and Analytical Framework

The central thesis of this essay is that cognitive outsourcing—the complete transfer of core cognitive tasks from learners to technology—constitutes the root generative mechanism of classroom risk. To establish this argument, it is essential first to distinguish two concepts that are frequently conflated: cognitive offloading and cognitive outsourcing. Cognitive offloading preserves the learner’s metacognitive control, whereas cognitive outsourcing involves the complete transfer of core cognitive tasks—such as judgment, reasoning, and metacognitive monitoring—to technology, resulting in the atrophy of cognitive engagement [6, 7].

I contend that the critical difference between the two lies not merely in degree but in kind: in offloading, the learner uses the tool; in outsourcing, the technology dominates the cognitive process. When human-technology interaction becomes asymmetrical, the tool ceases to be a transparent assistant and instead reshapes the epistemological standard of what counts as knowledge. This distinction is fundamental because it reframes the problem from one of excessive use to one of epistemic displacement—the quiet substitution of algorithmic authority for reasoned judgment.

Drawing on Cognitive Load Theory [8] and a human-technology interaction perspective on asymmetric mediation (Dwivedi et al., 2023), the paper traces four interlocking traps across thinking, subject, knowledge, and practice dimensions, proposing corresponding mitigation strategies through an integrated teaching-learning-assessment framework [9, 10].

2. The Generative Logic of Cognitive Outsourcing

2.1. From Cognitive offloading to Cognitive Outsourcing

Cognitive offloading begins as a neutral strategy to manage limited working memory. When students use tools to reduce extraneous load, the process is educationally productive [8]. However, when human-technology interaction becomes asymmetrical, the instrument ceases to be a transparent assistant and instead reshapes what counts as knowledge (Dwivedi et al., 2023).

The classroom manifestation of this transformation is gradual and concealed. Initially, students use AI as a tool to verify their own ideas; subsequently, they consult it as their primary or even sole authority; ultimately, the large language model becomes their default judge of what is true. In this process, learners slide from “using tools to think” to “letting tools think for them”, and cognitive offloading thereby degenerates into cognitive outsourcing. The Action Plan’s call for “improving students’ ability to think cognitively and solve complex problems” becomes particularly pertinent here. If core thinking processes are outsourced, this policy goal cannot be achieved. What makes this transformation particularly insidious is its self-reinforcing character. As learners outsource more cognitive labour, their capacity for independent reasoning atrophies, making them increasingly dependent on algorithmic assistance. This creates a dependency loop wherein the very tools intended to enhance learning progressively undermine the conditions for genuine intellectual development.

2.2. The Generative Logic of Cognitive Outsourcing

The first stage is a quiet shift in who owns the thinking. Students consult AI as their first and final authority rather than as a tool to check their own ideas. Learners often struggle to articulate their knowledge gaps when interacting with AI chatbots, fostering passive reliance. When AI’s response speed exceeds teacher feedback, the help-seeking target shifts—reconstructing trust from human to algorithm (Martens et al., 2023).

The second stage is a slow loss of intellectual stamina. Students trade short-term performance for long-term cognitive independence, accumulating a cognitive debt [5]. This debt reveals itself when learners face open-ended questions requiring independent judgment. Frequent AI use is negatively correlated with critical thinking, mediated by cognitive offloading [11]. An MIT study coined “cognitive debt” to describe how constant AI reliance impairs independent thinking processes.

The third and most troubling stage is the reconfiguration of what learners consider "good enough" knowledge. When algorithmic outputs consistently satisfy immediate task demands with minimal cognitive investment, learners' epistemic standards gradually erode. They become habituated to accepting syntactically coherent but conceptually shallow explanations, losing the capacity to recognise when understanding is genuinely inadequate. This epistemic satisficing—settling for minimally adequate explanations without critical evaluation—represents perhaps the most profound educational consequence of cognitive outsourcing.

3. Four-Dimensional Risk Analysis

3.1. Thinking Dimension: The Black-Box Trap

Algorithmic opacity and data bias reduce instruction from thinking cultivation to answer retrieval. Students bypass the constructive and interactive engagement essential to deep learning [12]. Teachers lose visibility into the reasoning that algorithmic fluency conceals. A student who studies diligently yet makes slow progress may be misjudged by evaluation algorithms as unengaged and inattentive.

Dialogue with the ICAP framework further illuminates this predicament. AI's generative responses often leave students at the interactive or even passive level of engagement, rather than the constructive or active level. When AI instantly generates complete answers, students' mental modelling process—the internally constructing of representations—is bypassed. They obtain the 'answer' but lose the crucial cognitive experience of "how the answer came to be".

Critically, the thinking dimension risk is not merely about the content of what students learn but about the cognitive processes they fail to develop. The black-box trap operates at two levels: the opacity of algorithmic reasoning and the opacity of students' own reasoning. This double opacity constitutes what I term "epistemic darkness"—a condition wherein neither the learner nor the teacher can reliably trace the provenance of knowledge claims, fundamentally undermining the transparency essential to educational relationships.

3.2. Subject Dimension: Alienation and Hollowed Interaction

Classroom interaction hollows out as algorithmic fluency replaces the hesitant, error-strewn processes through which understanding is socially forged [13]. Students no longer pose half-baked questions—these questions are precisely the authentic starting points of learning—but instead present AI-polished perfect questions. The system's instant responsiveness encroaches upon the teacher's professional judgement and the learner's epistemic agency, triggering a form of subject alienation [10]. In group discussions, AI's standard answers prematurely terminate the collision of viewpoints, and dissent, as a crucial mechanism of deep learning, is dissolved. The dialogic space that Mercer advocates thus loses its generative foundation.

In a randomised experiment with 117 students writing essays, found that while the ChatGPT group produced the greatest essay quality, participants did not learn more about the material, feel more motivated, or sustain greater interest compared to other groups [14]. In fact, students with only a checklist showed the highest engagement and time spent reviewing their work. The researchers named this pattern "metacognitive laziness"—learners offloaded the monitoring and evaluation of their own thinking to AI. The product improved; the learner did not.

3.3. Knowledge Dimension: Fragmentation and False Integrity

In the knowledge dimension, the core risk is false integrity: learners mistake AI-generated, syntactically coherent but conceptually fragmented outputs for genuine, structured understanding. The convenience effect lowers the threshold for epistemic satisficing. According to the ICAP framework, deep learning requires active integration of new information into existing knowledge networks [12]. However, AI's fluent outputs satisfy surface coherence while lacking deep inferential links, leading learners to engage in epistemic satisficing—accepting minimally adequate explanations without critical evaluation [11]. Because this illusion of comprehension is self-concealing, students

fail to detect their own conceptual gaps, particularly in ill--structured domains such as engineering design, where understanding requires mapping causal relationships across system components. Without metacognitive monitoring to distinguish informational fluency from structural coherence, learners accumulate shallow knowledge that cannot support transfer or problem--solving.

I argue that false integrity is not simply an error but a systematic feature of how AI-generated content interacts with human cognition. Humans are cognitively predisposed to accept fluent, coherent explanations as valid—a heuristic that generally serves us well in face-to-face interaction with other humans. AI exploits this heuristic without the corrective mechanisms of genuine dialogue, wherein interlocutors can probe, challenge, and elaborate. The result is a simulation of understanding that satisfies immediate assessment demands while foreclosing the generative struggle through which genuine expertise develops.

3.4. Practice Dimension: Disruption of Productive Ambiguity

The concept of productive ambiguity has deep roots in both Western and Chinese educational traditions. Dewey's notion of "problematic situations" and Vygotsky's "zone of proximal development" both emphasise that learning occurs at the edge of current competence, where the path forward is not immediately clear. In the Chinese tradition, the Confucian emphasis on "learning through reflection" similarly values the generative potential of uncertainty. From the perspective of Cognitive Load Theory [15], this productive ambiguity is not a design flaw but a necessary source of germane cognitive load—the mental effort dedicated to schema construction and deep understanding. When learners wrestle with incomplete information and tentative solutions, they actively build cognitive schemas that support transfer and problem-solving.

The instant responsiveness of AI, however, dissolves this productive ambiguity. By providing clear, immediate answers, AI reduces extraneous load but also strips away the very conditions that generate germane load. Learners are spared the struggle, but denied the growth that struggle produces. In the context of China's technology education reform, which privileges hands-on engineering practice [16], such instant smoothness displaces the very embodied inquiry the policy seeks to cultivate [17].

4. Mitigation Strategies: A Teaching-Learning-Assessment Framework

The rationality of the above framework is rooted in the principle of consistency [9]. Teaching, learning, and assessment constitute a dynamically calibrated closed-loop system: instruction provides cognitive scaffolds, learning generates process-oriented evidence, and evaluative feedback reciprocally informs design and pathways. This resonates with the UK DfE's 2026 progressive disclosure approach, where AI tool intensity is adjusted according to learner competence rather than substituted all at once.

I argue that effective mitigation requires not merely reducing AI use but fundamentally reconceiving the relationship between humans and algorithms in educational settings. The goal is to restore what I term "epistemic sovereignty": the learner's capacity for autonomous reasoning and the teacher's professional judgment as the organising principle of educational activity. Technologies should serve pedagogical purposes, not determine them.

4.1. Teaching: Rebuilding Classroom Thinking Space Through Generative Dialogue

Teachers should ask students to think and attempt solutions independently before introducing AI tools. More critically, algorithmic outputs should be transformed into objects of collective classroom discussion rather than endpoints. For example, teachers can design 'AI counter-argument tasks'—students first form independent viewpoints, then use AI to generate counter-arguments, and students must respond to these rebuttals. Algorithmic outputs are then turned into objects of collective classroom discussion rather than endpoints, rebuilding the generative interaction that Mercer [13] terms dialogic space.

The key principle here is temporal priority: independent reasoning must precede algorithmic consultation. This is not a matter of "resisting" technology but of establishing a cognitive order that privileges human judgment as the primary epistemic authority. A particularly instructive model comes from Felix Muzny's classroom at Northeastern University. Rather than discouraging AI-generated summaries, Muzny generates one himself using the same prompt, then asks students to critique what the model missed—misread figures, omitted nuances, misleading framings. The activity unfolds in class, in conversation, because students notice different things, and the discussion that follows is where learning actually happens.

4.2. Learning: Prioritising Independent Reasoning Over Algorithmic Acceptance

Teachers guide learners in reconstructing fragmented algorithmic outputs into coherent knowledge networks, reclaiming epistemic responsibility through structured integration tasks [8]. Specifically, an "AI Output Critique Sheet" can be introduced, requiring students to identify logical loopholes, knowledge blind spots, and value biases in AI responses, and to supplement their own evidence. This independent reasoning anchors higher-order goals and prevents surface drift [12]. However, the critique sheet must foster genuine critical engagement rather than superficial scepticism—mere rejection of AI answers without substantive reasoning is equally unproductive. Through this process of systematic critique and knowledge reconstruction, learners gradually rebuild the epistemological sophistication that false integrity has undermined.

A concrete instantiation can be found in a single classroom session. Students first generate an AI response to an open-ended question and save their output. In class, the teacher distributes three versions of the same prompt framed differently—for example, "explain the causes of the French Revolution" to a peer, to a child, and as a historiographical debate. Students compare these outputs in pairs, annotating shifts in framing, omissions, and which version offers the most nuanced account. The teacher then leads a discussion that asks not "which is correct?" but "what assumptions did the AI make, and what does that tell us about how to ask better questions?" Students conclude by revising their original AI-generated answer, justifying in one sentence what they changed and why. This sequence—compare, annotate, discuss, revise—transforms students from passive recipients of algorithmic output into active interrogators of its limitations within a single hour.

4.3. Assessment: Recalibrating Evaluation Around the Problem-Solving Process

The PDSA (Plan-Do-Study-Act) precision learning cycle can be meaningfully applied to the assessment dimension of the proposed framework. In the Plan phase, evaluation criteria should shift from judging final products to tracking learning processes. Drafts, error logs, and revision histories become central assessment objects [4]. During the Do phase, they collect these artefacts throughout a learning unit and document students' help-seeking behaviours. In the Study phase, teachers analyse the relationship between process indicators and genuine understanding, identifying early signs of cognitive outsourcing. Finally, in the Act phase, they refine the assessment rubric accordingly—for example, by giving greater weight to "productive struggle" indicators or adding scaffolding tasks for over-reliant students. This PDSA-driven approach transforms assessment from a static summative judgement into a formative tool.

The ultimate test of whether assessment has succeeded is not whether it produces a rich portfolio of documentation, but whether it changes how students relate to their own thinking. This suggests that process-trace assessment serves a higher purpose than diagnosis. Its real value lies in cultivating what might be called epistemic self-awareness—the habit of asking oneself: did I really understand this, or did I simply outsource the understanding? When students know that their reasoning path will be examined alongside their final answer, they develop a stake in the quality of that path itself. In this sense, the most successful assessment of cognitive outsourcing is when students no longer need to be asked to show their thinking because they have come to value showing it to themselves.

5. Conclusion

5.1. Summary of the argument

This paper has argued that cognitive outsourcing constitutes the root generative mechanism of classroom risk in digital-intelligent education. The distinction between cognitive offloading and cognitive outsourcing is not merely semantic. Cognitive offloading preserves the learner's metacognitive control, whereas cognitive outsourcing forfeits it entirely to algorithmic systems. This qualitative shift, I have contended, operates across four mutually reinforcing dimensions. The thinking dimension reveals how algorithmic opacity and data bias reduce instruction to answer retrieval, bypassing the constructive engagement essential to deep learning. The subject dimension exposes how the quasi-subjectivity of intelligent systems encroaches upon learner agency, hollowing out authentic classroom interaction by replacing genuine dialogue with algorithmic fluency. The knowledge dimension demonstrates how fragmented algorithmic outputs are mistaken for systematic understanding, generating false integrity that conceals genuine conceptual gaps. The practice dimension shows how instant responsiveness dissolves productive ambiguity, stripping away the generative friction through which understanding is forged. These four dimensions do not operate in isolation. Together they constitute a syndrome of epistemic displacement, the progressive substitution of algorithmic authority for reasoned judgment across the entire learning process.

The proposed teaching-learning-assessment framework responds not by rejecting AI but by restoring epistemic sovereignty. By epistemic sovereignty I mean the learner's capacity for autonomous reasoning and the teacher's professional judgment as the organising principle of educational activity. Teaching must establish temporal priority, requiring independent reasoning before algorithmic consultation and transforming AI outputs into objects of collective dialogue. Learning must move students from passive consumption to comparative critique and reconstruction, reclaiming the cognitive labour that outsourcing erodes. Assessment must make thinking visible through process-trace evaluation, cultivating the metacognitive awareness that outsourcing most directly threatens. In doing so, assessment can eventually render the external requirement to document redundant as students internalise the habit of self-monitoring.

Ultimately, the question is not whether to integrate AI into education, since that trajectory is irreversible. The question is whether we can do so without forfeiting the distinctly human capacities that education exists to cultivate. An affirmative answer requires not blanket resistance but vigilant calibration: a pedagogy that uses AI not as a cognitive prosthesis that replaces thinking, but as a provocateur that demands it.

5.2. Limitations and Future Directions

This conceptual analysis is necessarily limited by its reliance on theoretical synthesis rather than classroom-based empirical evidence. The four-dimensional framework, while grounded in established theories of cognitive load, dialogic pedagogy, and epistemic development, awaits empirical validation through longitudinal intervention studies across diverse educational contexts. Future research should test the proposed teaching-learning-assessment framework in real classrooms, examining whether the strategies described here actually reduce cognitive outsourcing and restore learner agency.

Future research can extend this work in three directions. First, longitudinal intervention studies should be conducted across diverse educational contexts to test the impact of the proposed framework on learners' cognitive engagement and critical thinking outcomes. Second, comparative research could explore differences in cognitive outsourcing patterns across subjects—for example, examining whether risks manifest differently in STEM disciplines versus humanities, where open-ended interpretation is central. Third, mixed-methods studies combining learning analytics with qualitative classroom observation could develop more nuanced indicators of cognitive outsourcing, enabling earlier and more accurate identification of at-risk learners. As AI technology continues to evolve, understanding and mitigating the hidden cognitive risks of integration will remain a core task for

educational research, one that requires ongoing dialogue between technical innovation and pedagogical principle.

References

- [1] Ministry of Education, Cyberspace Administration of China, Ministry of Industry and Information Technology, Ministry of Public Security, Ministry of Human Resources and Social Security, & China Disabled Persons' Federation. (2026). "AI+Education" Action Plan (Document No. Jiaokexin [2026] 1). Beijing: Ministry of Education.
- [2] Department for Education. (2026). Generative AI: product safety standards. Updated 19 January 2026. <https://www.gov.uk/government/publications/generative-ai-product-safety-standards>
- [3] S. Akgun, C. Greenhow. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2, 431–440. <https://doi.org/10.1007/s43681-021-00096-7>
- [4] Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, 6(1), 1–10.
- [5] Marsh, E. J., & Rajaram, S. (2020). The digital expansion of the mind: Implications of cognitive offloading for memory and learning. *Psychonomic Bulletin & Review*, 27(5), 1031–1038.
- [6] Wei, Enyuan. (2025). Educational traps of cognitive outsourcing in the digital-intelligent era and pathways to transcendence. *Open Education Research*, 31(1), 53–60.
- [7] Shaofeng Wang & Hao Zhang.(2026). Digital mentor paradox: Transforming algorithmic anxiety into professional growth through AI-scaffolded pedagogical reasoning and identity. *Teaching and Teacher Education*,176,105526.
- [8] Paas, F., & Sweller, J. (2020). Cognitive load theory: Advances in research on worked examples, affordances, and effort. *Educational Psychology Review*, 32(3), 681–701.
- [9] Cui, Y. (2024). Teaching-Learning-Assessment consistency: The key to deepening curriculum and teaching reform. *China Basic Education*, (1), 18–22.
- [10] Yu, S. & Wang (2018). The future role of artificial intelligence teachers. *Open Education Research*, 24(1), 16–28.
- [11] Michael Gerlich.(2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. *Societies*,15(1),6.
- [12] M. T. H. Chi, R. Wylie. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- [13] Mercer, N. (2022). Revisiting classroom talk: A communicational perspective. *Educational Psychologist*, 57(3), 239–252.
- [14] YizhouFan, LuzhenTang, HuixiaoLe, KejieShen, ShufangTan, YueyingZhao... & DraganGašević.(2024). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*,56(2),489–530.
- [15] Tamara van Gog, Fred Paas & John Sweller.(2010). Cognitive Load Theory: Advances in Research on Worked Examples, Animations, and Cognitive Load Measurement. *Educational Psychology Review*,22(4),375–378.
- [16] Zhang, Y., & Zheng, S. (2025). The real dilemmas and future pathways of teacher-student emotional reciprocity in the digital age: From the perspective of Goffman's dramaturgy. *E-education Research*, 46(4), 26–32.
- [17] Huang, R., Zhang, G., & Liu, M. (2024). Technical ethics orientation and risk regulation for smart education. *Modern Educational Technology*, 34(2), 13–22.
- [18] Juan Garzón, Eddy Patiño & Camilo Marulanda.(2025). Systematic Review of Artificial Intelligence in Education: Trends, Benefits, and Challenges. *Multimodal Technologies and Interaction*,9(8),84.
- [19] Sinha, T., & Kapur, M. (2021). When failure succeeds: Analyzing productive failure in an international context. *Journal of Educational Psychology*, 113(5), 1003–1019.

- [20] 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, Article 1111.
- [21] Wang, T., & An, J. (2026). The realistic representation, educational hidden worries and regulatory paths of cognitive outsourcing education in the intelligent era. *Journal of Southwest University (Social Sciences Edition)*, 52(1), 227–237, 291.
- [22] Zhong, B. (2021). Evaluation orientation of high-quality higher education system in the new era. *China Higher Education*, (1).