Application of AI in education and its influence on students' critical thinking

Ruimeng Zhang*

B.A. Education, University of Birmingham, Edgbaston, Birmingham, UK

Abstract. The deep integration of generative AI into education has attracted widespread attention, but existing research lacks a systematic exploration of the long-term influence mechanism of critical thinking. This article analyzes the educational reconstruction and its dual effects triggered by the evolution of AI from a "tool" to a "cognitive partner". Research shows that AI expands thinking training through personalized learning and immediate feedback. However, its algorithmic rationality implies cultural biases and a tendency towards data simplification, which may lead to the transfer of cognitive sovereignty and the compression of innovative thinking. The optimization strategies include: constructing a "technical buffer zone" to protect independent thinking; enhance algorithm transparency and cultural inclusiveness; promoting the transformation of teachers and improving the human-machine collaborative evaluation framework. The research emphasizes that AI education needs to strike a balance between technological empowerment and humanistic reflection to safeguard human cognitive sovereignty.

1 Introduction

With the explosion of Generative AI, education is undergoing its most profound paradigm shift since the digital revolution. According to UNESCO's 2023 Guidelines for the Application of Generative Artificial Intelligence in Education and Research, most countries around the world have formally incorporated AI tools into their basic education systems. The UK Digital Strategy, published on 13 June 2022, has been updated by the Department of Technology and Digital Economy to include digital skills and talent as one of six key areas. Such technological penetration not only improves teaching efficiency but also raises the alarm in the educational community: Existing research shows that students who over-rely on AI writing assistance are less able to demonstrate rigor and identify counterexamples [1]. This contradiction highlights the core blind spot of current educational technology research--existing literature mostly focuses on AI's optimization of knowledge transfer efficiency, but rarely systematically discusses its impact on students' higher-order thinking ability [2]. In particular, the long-term impact mechanism of Critical Thinking.

Critical thinking, as a key dimension of core literacy in the 21st century, is facing structural challenges brought by technological intervention. This study attempts to break through the

* Corresponding author: rxz309@student.bham.ac.uk

limitations of existing technology effectiveness studies and demonstrate a key proposition through the intersection of educational philosophy and cognitive science: when AI evolves from "teaching tool" to "cognitive partner", educators must reconstruct the instructional design paradigm and establish a dynamic balance between technology empowerment and thinking blank. This exploration has urgent practical significance for improving the ethical framework of AI education, optimizing teachers' Technological Pedagogical Content Knowledge (TPACK), and preventing the risk of technology alienation.

2 AI-driven Reconstruction of Educational Scenarios: Technological Penetration and Paradigm Transformation

2.1 AI technology penetration in educational scenarios: core applications and functional features

The penetration of artificial intelligence technology in educational scenarios has evolved from single tool assistance to systematic cognitive reconstruction. This technological intervention presents distinct hierarchical characteristics: At the surface functional level, the algorithm system optimizes the efficiency of traditional teaching links through automated processes, such as the immediate feedback of the intelligent marking system on standardized homework, or the recommendation of personalized learning paths driven by knowledge graphs. Such applications essentially continue the technological rational logic of the industrial age, converting quantifiable behavioral data in the educational process (such as the correct answer rate and learning duration) into fuel for algorithm optimization. Its core value lies in releasing teachers' repetitive labor and freeing up time and space resources for advanced teaching activities [3]. However, this efficiency-first technical positioning often falls into the paradox of self-reinforcement - when the system continuously iterates to improve the prediction accuracy, its underlying logic becomes increasingly closed, and eventually may compress the complex educational practice into a game of data fitting.

Deeper technological penetration occurs in the dimension of cognitive interaction. The adaptive learning system dynamically adjusts the presentation mode of teaching content and cognitive load intensity by monitoring students' cognitive trajectories in real time (such as the hesitation duration of problem-solving steps and the construction mode of knowledge associations). Such systems have transcended the physical boundaries of traditional teaching AIDS and evolved into "cognitive interfaces" that intervene in the thinking process. Their technical characteristics are manifested as algorithmic analysis and response to the learner's intention structure. For instance, writing AIDS not only corrects grammatical errors but also reveals logical breaks through the visualization technology of argumentative structures, externalizing the originally implicit thinking process into actionable improvement suggestions [4]. This deep intervention has sparked controversy at the ontological level of education: When algorithmic systems can simulate or even predict the cognitive paths of learners, the boundaries of educational time and space begin to dissolve, and the traditional dividing line between knowledge transmission and thinking training becomes blurred.

The deep-seated contradiction of technological penetration lies in its inherent tendency toward functional alienation. Most educational AI systems are built on the cognitive framework of Western centrism, and the default thinking paradigms of their algorithm models (such as linear logic and individual competition-oriented) have implicit conflicts with non-Western educational traditions [5]. For instance, a writing assessment system trained on English corpora might regard circuitousness in Eastern culture as a logical flaw. This dislocation of cultural cognition exposes the vulnerability of the claim of technological universality. The more fundamental predicament stems from the tension between algorithmic rationality and the humanistic nature of education - as the system continuously approaches the "optimal learning path" through behavioral data modeling, its technical essence is reshaping the value scale of education: quantifiable, predictable and controllable cognitive behaviors are given higher weights, while difficult-to-model human traits such as intuition and creativity are at risk of being marginalized. This redefinition of the essence of education by technological rationality constitutes the most concealed and profound transformation in the process of AI's deep penetration.

2.2 Reconstruction of the educational paradigm: from tool-assisted to cognitive synergy

The reconstruction of the educational paradigm by artificial intelligence is breaking through the traditional framework of instrumental rationality and moving towards a new ecosystem of human-machine cognitive collaboration. The driving force behind this transformation stems from the ontological shift triggered by technological intervention [6]. When algorithmic systems evolve from external tools to "thinking partners" embedded in cognitive processes, the spatio-temporal structure of education and the relationship between the subject undergo fundamental disintegration. The demarcated "teaching" and "learning" in traditional classrooms have been deconstructed into a dynamic cognitive network, and the consciousness flow of learners and algorithmic decision-making form a continuously interconstructing feedback loop. For example, in a mixed reality environment, students' real-time interaction with virtual agents through gestures not only acquires knowledge but also reshapes spatial perception and meaning construction patterns in technology-mediated embodied experiences. This deep collaboration blurs the boundary between biological intelligence and artificial systems, forcing educators to re-examine the core proposition of "what makes us human".

This reconstruction process implies profound practical paradoxes. On the one hand, the algorithm system provides a cognitive scaffold that is difficult for traditional education to reach for the cultivation of critical thinking through immediate feedback and personalized adaptation [7]. Students can simultaneously encounter the challenge of multicultural positions in virtual debates or test the completeness of hypotheses in real time in solving complex problems. On the other hand, the deep involvement of technology may lead to the implicit transfer of educational sovereignty--when the learning paths recommended by algorithms become increasingly accurate, students' motivation for independent exploration may be dissipated; When writing AIDS preset argumentation templates, the growth space for innovative thinking is facing compression. More crucially, the lag of the evaluation system exacerbates this tension: Standardized tests are difficult to capture the unexpected learning outcomes generated by human-machine collaboration, while process data may simplify thinking growth into quantifiable behavioral indicators.

Solving these paradoxes requires re-anchoring the value coordinates of the educational paradigm. Teachers must elevate themselves from the executors of technological applications to the designers of cognitive ecosystems. In their teaching designs, they should reserve "thought retention zones" that algorithms cannot reach [8]. For instance, they should stipulate that certain topics must be explored through original discussions without technological intervention, thereby safeguarding the irreplaceability of human intuition and ethical judgment. Educational researchers need to develop new assessment frameworks to incorporate cognitive synergy effects under technological intervention (such as the complementarity of human-machine argumentation and the ability to identify algorithmic biases) into the core literacy index system. These efforts all point to a fundamental principle: In the future of education where technology is deeply penetrating, humanity must maintain ultimate control over cognitive sovereignty--not by excluding technology, but by constantly reinterpreting and expanding the boundaries of thinking ability in the creative tension with technology.

3 Theoretical Framework of Critical Thinking Development

3.1 The core dimension of critical thinking

The essential dimension of critical thinking is always in the intersection of educational philosophy and cognitive science, and its theoretical construction reflects the deep vein of human rational self-reflection. The six-dimensional framework (interpretation, analysis, assessment, reasoning, interpretation and self-regulation) proposed in the existing relevant research report is essentially a modern transformation of Aristotelian logic tradition and Dewey's reflective thinking theory, the core of which is to establish the inter-subjective operability of rational criteria in thinking activities [9]. However, this classic model faces an ontological challenge in the digital age - when technological systems are involved in cognitive processes, the traditional dimensions of "analysis" and "evaluation" are reinterpreted: The object of analysis extends from the proposition itself to the implicit premise of algorithmic decision making, and the scope of evaluation extends from logical validity to the ethical consequences of technical tools.

The metacognitive dimension of critical thinking is often simplified as "self-regulation", which implies a deeper epistemological turn. The "Elements of thought" theory divides them into eight interacting elements: "purpose, problem, information, concept, hypothesis, perspective, reasoning, and implication" [10]. The essence of this theory is to build a self-referential system of thinking activities--learners should not only examine external propositions, but also continuously monitor the limitations of the internal cognitive framework. This reflexivity derives a new connotation in the educational scenario of technological intervention: when AI tools provide immediate feedback, the object of metacognition extends from individual thought processes to human-computer collaborative cognitive networks, requiring learners to simultaneously assess their reasoning deviations from the technological system. Therefore, the contemporary reconstruction of the dimension of critical thinking is essentially the practice of educational philosophy's resistance to technological universalism.

3.2 Cultivation path in traditional education

The traditional education system develops students' critical thinking through a variety of classical methods. The socratic method guides students to independently discover logical loopholes with continuous questioning. For example, in philosophy classes, teachers force students to deepen their thinking in continuous self-correction through progressive questions such as "What do you think justice is?" and "Is there an exception to this definition?" [11]. On the other hand, argumentative teaching trains the ability of multi-perspective analysis through role position conversion. For example, students are required to speak on behalf of opposing interests at random in simulated parliament, thus breaking through the inherent cognitive framework. Project-based learning (PBL) integrates thinking training into real problem solving. For example, when designing community environmental protection solutions, students need to independently collect contradictory data, coordinate multiple needs, and exercise systematic reasoning skills in complex situations. What these approaches have in common is the creation of cognitive conflict environments that shape habits of mind through practical engagement rather than passive acceptance.

However, the current education system faces structural constraints in the cultivation of thinking. The standardized assessment system forces teaching to focus on quantifiable knowledge acquisition. For example, the history curriculum emphasizes event memory rather than causal analysis, resulting in the Socratic questioning often reduced to the examination of the test site. Debate activities are mostly limited to extra-curricular interest groups, and ordinary classes are still dominated by one-way teaching by teachers, lacking normal critical thinking training. The implementation of PBL is limited by the uneven distribution of resources, and schools in underdeveloped areas often simplify projects for theoretical reasoning due to the shortage of equipment and teachers, losing the core value of practical reflection [12]. The more fundamental contradiction is that the progressive nature of critical thinking conflicts with the educational system's demands for rapid, interdisciplinary training - students need continuous training across disciplines, but the existing curriculum systems have fragmented and embedded thinking skills in each subject, making it difficult to form a coherent path to ability development.

4 Impact Analysis: AI and Critical Thinking

4.1 Positive influence

The integration of artificial intelligence technology in the educational scene is reconstructing the possibility boundary of critical thinking training from the cognitive dimension. Through algorithm-driven personalized learning path design, AI tools can break through the limitations of homogenization in traditional classrooms and provide a suitable thinking training rhythm for learners with different cognitive styles. This technical feature is closely integrated with the self-regulating dimension of critical thinking [9]. When the system dynamically adjusts the complexity of the problem to match the learner's most recent development zone, individuals can enhance their analytical and reasoning abilities in the continuous iteration of cognitive conflicts. From the perspective of educational philosophy, this kind of technological intervention can be seen as a digital extension of Vygotsky's scaffolding theory, whose core value is to liberate teachers from standardized knowledge transfer and instead focus on the role reconstruction of higher-order thinking facilitators.

The breakthrough development of generative technologies has injected new paradigms into critical thinking. The adversarial arguments generated by large language models force learners to constantly test the completeness of their arguments between multiple positions. The "artificial cognitive crisis" created by this technology essentially continues the dialectical spirit of Socratic interrogation, but achieves a scale and immediacy that traditional education cannot achieve through algorithms [13]. The digital transformation of teacher professional development further amplifies the positive impact of AI. The real-time analysis of classroom interactions by intelligent systems provides educators with cognitive behavioral mapping at the micro level, which enables the practical application of Bloom's classification from abstract theory to precise operation. When teachers can identify implicit logical faults in students' arguments through algorithms, their teaching intervention can go beyond empirical intuition and turn to evidence-based precision guidance.

4.2 Potential risks and challenges

The widespread application of AI educational tools is reshaping the underlying logic of teaching practice and causing new structural contradictions. Teachers are generally faced with the problem of controlling the degree of tool involvement: when the writing aid system can point out the logical loopholes immediately, students may skip the self-elaboration and directly adopt suggestions, resulting in in-depth argumentation training that is somewhat superficial. Although the adversarial arguments generated by AI in classroom debate activities can expand the breadth of thinking, the cultural bias of the default algorithm model may compress the multi-cognitive paradigm into a single technical criterion; for example, the contradiction reconciliation in eastern dialectical thinking is regarded as a logical defect. The digital transformation of assessment systems is exacerbating the conflict between standardization and creativity: the quantitative preference of automated scoring systems for argumentative structures is forcing teachers to adapt instructional designs to machine-recognizable templates, and this reverse regulation is eroding the openness that should be associated with critical thinking training [14].

The technology dependence of educational resources has formed a new ability gap. Due to the lack of algorithm interpretation ability, some teachers find it difficult to guide students to dialectically use AI tools, and instead transfer the technical decision-making power to the system preset process. When accessing intelligent education platforms in less developed regions, they are often caught in a dilemma due to inadequate cultural adaptation - the use of Western-centric algorithm models will distort local cognitive traditions, and the cost of customized development is beyond the actual affordability [15]. A more insidious risk lies in the distortion of cognitive development monitoring: the trajectory of thought growth that AI systems chart through behavioral data may reduce complex critical abilities to quantifiable metrics (such as number of refutations or complexity of argument structures), ignoring core dimensions that are not computable, such as intuition and ethical judgments. These practical dilemmas together point to the fundamental challenge in the digital transformation of education: how to protect the light in human thinking activities that cannot be replaced by algorithms while improving the efficiency of technology empowerment.

5 Optimization Strategies and Future Recommendations

5.1 Principles of educational design

With the deep integration of artificial intelligence (AI) into educational settings, instructional design must reimagine the tripartite relationship among teachers, students, and technology [16]. This transformation necessitates a paradigm shift in pedagogical roles: teachers are no longer confined to the traditional role of knowledge transmitters but must evolve into cognitive curators--a term denoting educators who critically evaluate and strategically integrate AI tools during lesson planning. As cognitive curators, teachers proactively delineate the functional boundaries of AI systems, ensuring alignment with pedagogical objectives while mitigating over-reliance on automation. To safeguard cognitive autonomy, curriculum frameworks should incorporate a technological intervention buffer zone--a structured phase in which students engage in independent problem-solving without AI assistance. For instance, in writing assignments, educators might implement a no-AI first-draft requirement, mandating that students complete initial drafts through unaided critical reflection before utilizing AI tools for refinement. This buffer zone serves dual purposes: it preserves the foundational space for original thought generation and reinforces metacognitive skills such as self-regulation and analytical reasoning.

At the systemic level, educational stakeholders must adopt ethical and inclusive principles to govern AI integration. This requires two complementary strategies: Capacity building for educators: Teachers should receive training in algorithmic literacy--the ability to critically assess the ethical implications and cultural assumptions embedded in AI tools [17]. Such training empowers educators to identify potential biases in data-driven systems and make informed pedagogical decisions. Educational technology providers should collaborate with schools to design adaptable AI solutions that align with diverse learning objectives. Rather than demanding full transparency of proprietary algorithms, this partnership could focus on co-developing pedagogically aligned intervention modes--for example, enabling educators to adjust the scope of AI support based on students' developmental stages. Additionally, interdisciplinary teams should implement bias stress-testing frameworks in curricula. A practical approach involves embedding critical reflection activities where students analyze how cultural or ideological biases in AI outputs (e.g., historical analysis tools) might skew interpretations. For under-resourced regions, educational technology should prioritize culturally sustaining design--developing tools that respect local knowledge systems instead of imposing standardized models. For instance, lightweight AI assistants could scaffold critical thinking through locally relevant problem-solving scenarios rather than replicating complex systems designed for technologically advanced contexts [18].

These principles collectively affirm a foundational imperative: As AI becomes ubiquitous in education, pedagogical sovereignty must remain rooted in human agency, ensuring technology amplifies rather than dictates the cultivation of critical self-awareness.

5.2 Direction of technological improvement

The optimization direction of artificial intelligence educational technology needs to revolve around the core goal of cultivating critical thinking and constructing a multi-party collaborative improvement framework. Technology developers should focus on the educational attributes of tools rather than merely functional expansion, and prioritize enhancing the transparency and interpretability of algorithms. By visualizing the reasoning logic (such as multipath analysis demonstrating the generation of arguments), the technical system can transform from an "answer provider" to a "mind mapper", enabling students to visually compare the similarities and differences between their logic and algorithmic decisions, thereby stimulating the awareness of questioning and reflection. Meanwhile, technical design needs to embed multicultural cognitive models to prevent tools from solidifying thinking paradigms due to a single training dataset. For instance, integrating the evaluation dimensions of Eastern dialectical thinking and Western deductive logic in language models can provide students with a practical field for cross-cultural criticism [19].

Educational institutions need to redefine the positioning of digital classrooms, shifting from technology-driven to a human-machine symbiotic hybrid model. Schools should establish a dynamic adjustment mechanism for AI intervention, and flexibly adjust the intensity of technological participation according to teaching objectives - utilize intelligent tools to broaden the boundaries of information in the knowledge integration stage, and limit automated assistance in the deep critical thinking stage to protect the space for independent thinking; The transformation of the teacher role is a key hub for technology empowerment. Educators need to go beyond the level of tool operators and become strategic designers of technology applications. By reverse designing teaching tasks (such as requiring students to manually trace and modify the logic after using AI writing assistance), teachers can guide students to deconstruct the implicit rules behind technical decisions. Meanwhile, by using AI to generate adversarial viewpoints (such as automatically generating arguments that are contrary to the text), teachers can create cognitive conflict situations, driving students to gradually build up the critical resilience to resist technical authorities through multi-source information comparison and logical calibration. The activation of students' subjectivity is the ultimate destination of technological optimization [20]. Educational design needs to strengthen metacognitive training and transform the process of technology usage into an opportunity for self-monitoring.

The ultimate goal of technological improvement is not to replace human thinking, but to build an educational ecosystem where "technology serves as a mirror" - the algorithm system acts as a developer in the thinking process, helping learners to more clearly observe the blind spots in their cognition. The realization of this goal requires developers, schools, teachers, and students to reach a value consensus: the intensity of technological intervention always serves the need for deepening thinking, while educational sovereignty always belongs to people's critical self-awareness. Only by maintaining a balance between instrumental rationality and humanistic reflection can artificial intelligence truly become a collaborator rather than a dissolver in cultivating higher-order thinking abilities.

6 Conclusion

The deep application of artificial intelligence educational tools is reshaping the basic paradigm of teaching practice, which brings both opportunities for efficiency improvement and worries about the dissolution of educational subjectivity. The core challenge of the current practice is to balance the dialectical relationship between technological empowerment and cognitive autonomy - expanding teachers need to develop new technological literacy, be good at using AI tools to expand the dimensions of thinking training, and be able to accurately identify the critical point of algorithm intervention. The educational practice of the future will always be in the tension between technological iteration and humanistic persistence, and its success will depend on whether teachers can establish a culture of "technological skepticism" in the classroom - not simply rejecting or embracing AI, but training students to become strategic critics of the use of technology. The development of this ability will ultimately require the education system to recognize a fundamental fact at the institutional level: true critical thinking always grows out of the collision of ideas between people, and technology can only be a participant in this dialogue, not a leader.

References

1. E. R. Sykes, AI-assisted academic writing: A comparative study of student-crafted and ChatGPT-enhanced critiques in ubiquitous computing. Int. J. Adv. Comput. Sci. Appl. 15 (2024).
2. O. Zawacki-Richter, V. I. Marín, M. Bond, F. Gouverneur, Systematic review of research on artificial intelligence applications in higher education–where are the educators? Int. J. Educ. Technol. High. Educ. **16**, 1 (2019).
3. I. A. P. Wogu, S. Misra, E. F. Olu-Owolabi, et al., Artificial intelligence, artificial teachers and the fate of learners in the 21st century education sector: Implications for theory and practice. Int. J. Pure Appl. Math. **119**, 2245 (2018).
4. A. Skulmowski, The cognitive architecture of digital externalization. Educ. Psychol. Rev. **35**, 101 (2023).

5. Y. Ofosu-Asare, Cognitive imperialism in artificial intelligence: counteracting bias with indigenous epistemologies. AI Soc. 1 (2024).
6. E. Creely, K. Janssen, Onto-epistemological understandings of generative artificial intelligence in education. Int. J. Chang. Educ. (2025).
7. F. Wei, Cultivation of students' critical thinking in the era of artificial intelligence–Based on the perspective of training mode and education subjects. In Appl. Intell. Syst. Multimodal Inf. Anal. **2**, 786 (2021).
8. V. E. S. Seaba, Revolutionizing education: Exploring the potential of AI-enabled brain-based learning for enhanced cognitive development. Open Access Libr. J. **10**, 1 (2023).
9. P. A. Facione, Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (California Academic Press, 1990).
10. R. W. Paul, Critical and reflective thinking: A philosophical perspective. In Dimens. Think. Cogn. Instr. 445 (2013).
11. A. J. Grondin, Effectiveness of the Socratic method: A comparative analysis of the historical and modern invocations of an educational method (2018).
12. K. R. Giva, S. E. Duma, Characteristics and critical success factors for implementing problem-based learning in a human resource-constrained country. Curationis **38**, 1 (2015).
13. J. Jasin, H. T. Ng, I. Atmosukarto, et al., The implementation of chatbot-mediated immediacy for synchronous communication in an online chemistry course. Educ. Inf. Technol. **28**, 10665 (2023).
14. L. Perla, V. Vinci, Enhancing authentic assessment in higher education: Leveraging digital transformation and artificial intelligence. AIxEDU 1 (2023).
15. H. Li, Multicultural data assistance mining analysis for ideological and political education in smart education platforms using artificial intelligence. Wirel. Netw. **31**, 567 (2025).
16. A. Rauf, S. Nadeem, L. Tahir, Integrating artificial intelligence into curriculum design. Multidiscip. J. Emerg. Needs Curric. **1**, 10 (2024).
17. M. Turilli, L. Floridi, The ethics of information transparency. Ethics Inf. Technol. **11**, 105 (2009).
18. G. Ladson-Billings, Toward a theory of culturally relevant pedagogy. Am. Educ. Res. J. **32**, 465 (1995).
19. M. Cui, N. A. B. A. Aziz, A. B. Hassan, Research on cross-cultural education communication platform based on artificial intelligence. In Proc. 2024 4th Int. Conf. Internet Technol. Educ. Informatization 265 (2024).
20. W. Li, J. Tong, Research on the problems and countermeasures of the imbalance of students' subjective status in the age of artificial intelligence. Asia Pac. Econ. Manag. Rev. 1 (2024).