



From Tool to Partner to Agent? A Critical Systematic Review of Human-AI Relationships in Literary Education

Mingqin Zhou 

Wannan Medical University, 22 Wenchang West Road, 241002 Wuhu, Anhui, P.R. China

19905508199@163.com

Abstract. LLMs are entering literary education, often framed as shifting from tool to partner to agent. Based on a PRISMA systematic review of 67 studies (2021–2026), we propose a revised Tool–Partner–Agent framework evaluated by technical maturity, pedagogy, and interpretive autonomy. Key findings: (1) 69% of studies remain at Tool phase—strong efficiency gains ($d=0.59-0.71$) but shallow understanding. (2) Partner phase enhances fluency ($d=0.58-0.71$) but reduces originality ($d=-0.43$) and authorial voice, with educational level as a strong moderator. (3) No mature Agent system exists; speculative designs risk algorithmic authority and the automation of interpretation paradox. We conclude with design principles and six research priorities.

Keywords: Human-AI collaboration; literary education; systematic review; human agency; critical AI literacy; creativity

1 Introduction

LLMs (GPT-4, Claude 3) offer text generation and dialogue, transforming education [1]. Yet applications concentrate in STEM and language learning, leaving humanities understudied [2]. Literary education—emphasizing close reading, ambiguity, and empathy—could benefit from AI (e.g., character analysis, reduced cognitive load) [3]. However, the human-AI relationship is unclear: AI is called tool, partner, or agent without empirical evaluation of risks to human agency, a core humanities value. This lack of clarity is particularly problematic in literary education, where the cultivation of independent interpretation, critical reflection, and emotional resonance lies at the heart of pedagogical goals. Reducing AI to a mere tool may overlook its generative potential, whereas prematurely positioning AI as a partner or agent risks displacing the learner’s interpretive authority. As Beck and Levine [4] warn, AI risks being “too correct, too fast,” offering polished answers that short-circuit the slow, uncertain, and deeply human process of making meaning from literature. No existing review provides a phase-based synthesis for literary education that systematically examines how different levels of AI autonomy affect human agency, creativity, and interpretive depth. To address this gap, this review: (1) proposes an agency-centered Tool–Partner–Agent framework; (2)

© The Author(s) 2026

C. F. Peng et al. (eds.), *Proceedings of the 2026 5th International Conference on Humanities, Wisdom Education and Service Management (HWESM 2026)*, Advances in Social Science, Education and Humanities Research 1024,

https://doi.org/10.2991/978-2-38476-593-5_86

synthesizes 67 studies; (3) identifies the automation of interpretation paradox; (4) offers design principles.

2 Methodology

Following PRISMA 2020, we searched Web of Science, Scopus, ERIC (Jan 2021–Mar 2026) with terms for AI, literary education, and human-AI interaction. Search strings combined keywords such as (“large language model*” OR “generative AI” OR “GPT” OR “LLM”) AND (“literature education” OR “literary reading” OR “close reading” OR “creative writing instruction”) AND (“human-AI collaboration” OR “AI assistant” OR “tool-partner-agent”). After deduplication, 1,247 records were screened by title and abstract against inclusion criteria: (a) peer-reviewed empirical studies or systematic reviews; (b) focusing on K–12 or higher education literary contexts; (c) involving LLM-based AI interaction. Full-text screening of 213 articles yielded 67 studies meeting all criteria. Quality was assessed using MMAT [5]; inter-coder reliability $\kappa=0.84$. Thematic synthesis and deductive coding categorized studies into Tool, Partner, and Agent phases based on operational definitions of AI autonomy and human agency. Two researchers independently coded all included studies, resolving disagreements through consensus discussion.

3 The Tool Phase: AI as an Efficient but Shallow Assistant

AI acts as a passive, user-controlled instrument [6]. Technologies include sentiment analysis, topic modeling, and translation. Applications: faster text analysis (e.g., 40% faster pattern identification) but no deeper comprehension [7]; 67% of students worry instant answers reduce motivation to infer meaning [3]. For L2 learners, syntactic accuracy improves ($d=0.68$) but not narrative voice [8]. Leveled texts boost motivation but risk literary quality. Critical assessment: clear efficiency gains but limited literary thinking value; must be paired with interpretive scaffolding.

4 The Partner Phase: Collaborative Co-Creation and Its Double-Edged Sword

In the Partner phase, AI acts as an interactive collaborator that sustains context, generates alternatives, and challenges user assumptions while retaining human agency [6,9]. Core technologies include generative LLMs (GPT-4, Claude 3), prompt engineering, and retrieval-augmented generation (RAG). Educational applications span collaborative writing (improving fluency and coherence but reducing experimentation and affecting authorship ownership), intelligent literary interpretation (simulating critical theories but yielding less emotionally resonant readings and losing cultural nuance), and adaptive teaching (providing personalized scaffolding yet risking stylistic innovation) [10–12].

A systematic review of 25 studies confirms that AI’s effects on writing and creativity are significantly moderated by educational level, usage patterns, and pedagogy [13]. As shown in Table 1, graduate students use AI more strategically than undergraduates, who in turn outperform high school students; elementary students gain fluency but risk formulaic writing.

Table 1. Effects by Educational Level

Educational Level	Observed Effects	Caveats
Elementary (K-5)	Improved narrative structure and fluency	Risk of formulaic writing
Middle School (6-8)	Enhanced descriptive richness	Reduced experimentation
High School (9-12)	Increased productivity	Lower originality in AI-heavy conditions
Undergraduate	Higher plot complexity, lower authorial voice	Automation bias prevalent
Graduate	More strategic, selective AI use	Better critical evaluation skills

The creativity paradox persists: appropriate AI use stimulates creativity, but overuse impairs originality. A quasi-experiment with 124 undergraduates found that AI collaboration enhanced plot complexity ($d = 0.71$) and descriptive richness ($d = 0.58$) but reduced narrative originality ($d = -0.43$) and authorial voice ($d = -0.39$) [13]. Competing evidence suggests AI can act as an instructional coach, improving writing quality more effectively than web searching with less effort. User-initiated collaboration yields stronger perceived ownership (the IKEA effect), while linguistic innovation peaks when AI acts as a provocateur [14]. However, high AI support correlates with skill atrophy in independent error identification, revealing a trade-off between scaffolding and autonomous ability [15].

Critical discussion highlights three concerns. First, the paradox of partnership: AI boosts fluency but risks stylistic homogenization via pattern recombination, and students often fail to recognize over-reliance [1,7]. Second, power asymmetry favors fast, fluent AI, leading to uncritical deference among less literate users (Singapore & Vietnam, 2024). Third, educational level moderates effects, making tiered scaffolding and explicit critical AI literacy essential [13]. Metacognitive scaffolds—prompting critique and modification of AI suggestions—promote positive outcomes; without them, AI reduces stylistic diversity and creative ownership [12].

Evidence-based design principles for Partner-phase implementation include: explicit metacognitive prompts documenting AI use decisions, comparative evaluation of AI vs. human-authored content, gradual autonomy progression from Tool to Partner phases, multiple-alternative generation to support comparative thinking, and post-collaboration reflection on individual creative contributions.

5 Comparative Analysis and Conceptual Framework

We propose a two-dimensional model based on AI autonomy (low/high) and human agency (low/high). The Tool phase (low AI autonomy, high human agency) is suitable for most tasks. The Delegation Zone (high AI autonomy, high human agency) is under-researched. The Agent phase (high AI autonomy, low human agency) is not recommended for literary education. As shown in Table 2.

Table 2. Comparative Analysis of Tool, Partner, and Agent Phases

Dimension	Tool Phase	Partner Phase	Agent Phase (Speculative)
AI role	Passive instrument	Interactive collaborator	Proactive decision-maker
Human agency	High (full control)	Medium-High (shared with scaffolds)	Low (AI initiates)
Typical tasks	Fact retrieval, basic analysis, translation	Co-writing, multi-perspective interpretation	Autonomous scaffolding, discussion moderation
Technical basis	Text mining, sentiment analysis, topic modeling	Generative LLMs, prompt engineering, RAG	Autonomous goal-setting, student modeling
Empirical evidence	Strong for efficiency ($d = 0.59-0.71$), weak for depth	Mixed: higher fluency ($d = 0.58-0.71$), lower originality ($d = -0.43$)	None in literary education
Educational level suitability	All levels with appropriate scaffolding	Undergraduate and above with explicit training	Not recommended
Primary risk	Shallow reading, answer-seeking	Stylistic homogenization, automation bias	Erosion of interpretive autonomy
Recommended use	Yes, with explicit interpretive follow-up	Yes, with metacognitive prompts and reflection	No, without further research and safeguards
Key competencies required	Basic digital literacy	Critical AI literacy, metacognitive skills	Not applicable

6 Conclusion

This review finds the Tool phase dominates (69%, efficiency gains only), the Partner phase boosts fluency but reduces originality and voice (moderated by education level), and the Agent phase is theoretical. These findings reveal a critical gap between AI's technical capabilities and its pedagogical integration in literary education. While LLMs

offer unprecedented efficiency in text processing and idea generation, their use in the Tool phase often results in shallow reading and answer-seeking behaviors rather than fostering interpretive depth. The Partner phase presents a double-edged sword: AI enhances narrative structure and descriptive richness, yet consistently reduces originality and authorial voice, with educational level as a significant moderator. Graduate students demonstrate more strategic and selective use, whereas younger learners are more vulnerable to automation bias and stylistic homogenization. We identify the automation of interpretation paradox—the phenomenon in which AI accelerates and smooths the interpretive process but simultaneously undermines the slow, uncertain, and personally meaningful engagement that lies at the heart of literary education. This paradox suggests that the very features making AI attractive (speed, fluency, correctness) may be antithetical to developing human agency in interpretation. AI must be a calibrated scaffold: Tool for low-stakes tasks (e.g., vocabulary support, basic plot recall), Partner with explicit metacognitive prompts and critical AI literacy scaffolds, and Agent not recommended without radical redesign and safeguards. Future research should prioritize longitudinal designs, develop validated instruments for measuring interpretive autonomy, and investigate instructional strategies that preserve originality and voice while leveraging AI's generative strengths. Practitioners should adopt tiered integration models that gradually increase AI autonomy only as students' critical evaluation skills mature.

References

1. Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S.: Paper review: 'Sparks of artificial general intelligence: Early experiments with GPT-4'. <https://cis.temple.edu/tagit/presentations/Sparks%20of%20AGI%20Early%20Experiments%20with%20GPT4.pdf>, last accessed 2026/04/29
2. Wang, P., Jing, Y., Shen, S.: A systematic literature review on the application of generative artificial intelligence (GAI) in teaching within higher education: Instructional contexts, process, and strategies. *Internet and Higher Education* **65**, 100996 (2025)
3. Xia, T., Pan, X., Cao, M., Guo, J.: An investigation of college students' acceptance of AI-assisted reading tools: An expansion of the TAM and SDT. *Education and Information Technologies* **30**(13), 18031–18058 (2025)
4. Beck, S.W., Levine, S.: The next word: A framework for imagining the benefits and harms of generative AI as a resource for learning to write. *Reading Research Quarterly* **59**(4), 706–715 (2024)
5. Hong, Q.N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.C., Vedel, I., Pluye, P.: The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *Education for Information* **34**(4), 285–291 (2018)
6. Flathmann, C., McNeese, N.J., Sengupta, S., Johnson, E.: Exploring trust, acceptance, and behavioral differences when humans collaborate with large language models as tools and teammates. *ACM Transactions on Interactive Intelligent Systems* **15**(4), 1–33 (2025)

7. Zhang, T., Zhang, P., Ma, H.: Tracing the topic evolution of AI literacy literature: From knowledge enlightenment to knowledge symbiosis. *Knowledge Organization* **52**(7), 44114 (2025)
8. AlTwijri, L., Alghizzi, T.M.: Investigating the integration of artificial intelligence in English as foreign language classes for enhancing learners' affective factors: A systematic review. *Heliyon* **10**(10), e31053 (2024)
9. Park, J.: A systematic literature review of generative artificial intelligence (GenAI) literacy in schools. *Computers and Education: Artificial Intelligence* **9**, 100487 (2025)
10. Mysore, S., Das, D., Cao, H., Sarrafzadeh, B.: Prototypical human-AI collaboration behaviors from LLM-Assisted writing in the wild. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Suzhou, China, 2025.
11. Bellot, A.R., Gutiérrez-Colón, M.: All assignments are equal, but some assignments are more equal than others: Exploring student responses to AI and peer-generated alternative endings in Orwell's animal farm. *Journal of Adolescent and Adult Literacy* **69**(4), e70025 (2026)
12. Kim, J., Lee, H., Cho, Y.H.: Learning design to support student-AI collaboration: Perspectives of leading teachers for AI in education. *Education and Information Technologies* **27**(5), 6069–6104 (2022)
13. Bal, M., Arseven, T.: Students' writing skills and creativity in the age of artificial intelligence: A systematic review of creativity across diverse AI-supported writing contexts. *Thinking Skills and Creativity* **60**, 102118 (2026)
14. Nordahl Fundal, H., Eide Rambøll, J., Olsen, K.: Alignment, exploration, and novelty in human-AI interaction. *arXiv e-prints*, arXiv: 2512.17117 (2025)
15. Cen, X., Shakibaei, G.: Enhancing creative writing through AI-powered co-creation with cognitive and emotional outcomes. *Sci Rep* **16**, 4407 (2026)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

