Vol. 24 (2026): Proceedings of the 2nd International Conference on Linguistics, Psychology and Creative Education (ICLPCE 2026)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/iclpce.20261419

# Human-AI Collaborative Learning: A Hybrid LLM Orchestration Framework Driven by Cognitive Dissonance

**Yiqing Wu***

*Shandong Normal University, Jinan, Shandong, China*

**Corresponding author: Yiqing Wu.*

## Abstract

The rapid evolution of generative AI, particularly large language models (LLMs), is reshaping education from individualized instruction toward collective intelligence-driven collaborative learning. However, current intelligent educational systems face critical challenges, including fragmented collaboration ecosystems, isolated AI functionalities, and superficial learning analytics that hinder deep interactions among students, peers, and instructors. To address these gaps, this review proposes an integrated theoretical–technological–analytical framework grounded in constructivist learning theory and cognitive dissonance theory, conceptualizing knowledge as socially co-constructed. The framework comprises three core innovations: (1) a multi-agent collaborative architecture for real-time interaction; (2) a hybrid LLM orchestration mechanism for dynamic model deployment; and (3) a data-driven analytical engine leveraging semantic NLP to identify cognitive conflicts and trace collective intelligence trajectories. The framework advances next-generation intelligent collaborative learning systems by bridging learning sciences, artificial intelligence, and educational data mining. It also identifies key implementation challenges, including technical stability, ethical data governance, and pedagogical adoption, and proposes future directions such as affective computing integration, cross-institutional learning ecosystems, and multidimensional assessment reform. This work contributes a theoretically grounded and operationalizable design paradigm for AI-supported collaborative learning systems.

### Keywords

hybrid large language models, crowd-intelligence collaborative learning, constructivism, intelligent educational system, learning analytics

## 1. Introduction

Artificial intelligence, especially the breakthrough development of large language models (LLMs), is reshaping educational paradigms by shifting learning from passive knowledge reception toward active, human-AI co-creation [1]. However, how to systematically support deep, multi-agent collaborative knowledge construction through AI remains an unresolved research problem [2]. Despite the increasing integration of AI features in contemporary online learning platforms, their application largely remains confined to individualized tutoring or static content generation, failing to provide systematic support for the multi-layered and evolving dynamics of authentic student-student-AI-teacher collaboration [3]. A key limitation is that current AI tools often operate peripherally to the core collaborative workflow; their functionalities are siloed, and their capacity to perform deep analysis of rich interaction data—particularly at the level of cognitive and social meaning, remains underdeveloped [4]. Consequently, pedagogical interventions struggle to target critical moments of knowledge construction, and learners miss out on meaningful feedback about their collaborative processes. To address these systemic shortcomings, this paper proposes a comprehensive, theory-grounded framework designed to transcend narrow, tool-centric approaches. It offers an integrated solution for intelligent collaborative learning environments by first establishing a theoretical foundation that synthesizes constructivist and cognitive dissonance theories to explain the social negotiation of knowledge and the role of cognitive conflict in driving deeper learning [5]. Building on this foundation, the paper then articulates a three-module architectural framework comprising a multi-agent collaborative ecosystem, a hybrid LLM orchestration mechanism, and a data-driven analytical engine, each designed to address specific identified limitations while synergistically reinforcing the others [6,7]. Finally, the discussion extends to the framework's potential applications, the significant implementation challenges it faces, and fruitful directions for future research. By bridging learning theory, AI technology, and educational data science, this work aims to catalyze the development of a new generation of learning systems that empower not only individual learners but the very process of collective intelligence itself.

## 2. Theoretical Background

### 2.1 Learning Theory: Social Constructivism as the Foundational Paradigm

The theoretical underpinning of the proposed framework is rooted in social constructivism, a dominant paradigm in learning sciences. This theory asserts that knowledge is not a commodity to be transmitted from teacher to student, but is actively constructed and negotiated by learners through social interaction and discourse within a specific cultural context [1]. In the context of intelligent learning environments, this shifts the role of technology from a passive delivery mechanism to an active participant in the social construction of knowledge. Consequently, Artificial Intelligence (AI) agents are conceptualized as cognitive partners or intelligent mediators [2]. Their primary function extends beyond information provision to include scaffolding collaborative dialogue, introducing contrasting viewpoints to stimulate critical thinking, and helping to synthesize diverse contributions from group members, thereby directly facilitating the collaborative knowledge-building process [3]. This theoretical commitment necessitates the design of learning systems that inherently support equitable, multi-directional, and real-time interaction among all human and artificial agents.

### 2.2 Cognitive Mechanism: Cognitive Dissonance as the Driver of Deep Learning

To operationalize and leverage the internal dynamics of collaborative learning, the framework integrates cognitive dissonance theory as its core psychological mechanism [4]. The theory posits that individuals experience a state of psychological discomfort, or dissonance, when they simultaneously hold two or more psychologically inconsistent cognitions (e.g., beliefs, attitudes, pieces of knowledge) [5]. This aversive state creates a motivational drive to reduce the inconsistency, often leading to changes in belief, attitude, or behavior. Within collaborative learning, such dissonance is not a dysfunction to be avoided but an essential and valuable catalyst. It naturally emerges from the encounter of divergent perspectives, conflicting interpretations, and debates over problem-solving strategies. An intelligent system informed by this framework is therefore designed to detect these critical moments of cognitive dissonance within group discourse, analyze their nature and potential, and strategically intervene to guide the group toward a constructive resolution. The goal is to transform epistemic conflict into a productive opportunity for deep reflection, perspective-taking, and sustainable conceptual change [6].

### 2.3 System Design Theory: A Layered Model for Human-AI Collaboration

To translate theoretical insights into system-level design principles, this study adopts a three-layer model of Human–AI collaboration. The translation of the above learning and cognitive theories into a concrete technical architecture is guided by a three-layer model of Human-AI collaboration, synthesized from human-computer interaction and hybrid intelligence literature [3]. This model conceptualizes the evolving roles of AI along a spectrum of agency and contextual integration: At the foundational level, the AI functions as a sophisticated tool, executing specific, predefined tasks (e.g., grammar checking, information retrieval) with high reliability but minimal contextual awareness or adaptive interaction. Moving to the intermediate level, the AI evolves into a co-learner or team member. It engages in contextually-aware, turn-taking dialogue, understands and contributes to shared task goals, and adapts its contributions dynamically, forming a synergistic partnership with human learners. Finally, at the most advanced level, the model envisions AI as an integral and adaptive component of a dynamic learning ecosystem. Here, AI, teachers, and students form a co-evolving system where the AI's behavior, roles, and supports are not static but are dynamically calibrated in response to long-term pedagogical objectives, the evolving social fabric of the group, and the collective learning trajectory [9].

This layered model serves as the principal design compass for the proposed framework, directly informing critical decisions on how AI agents should behave, adapt their roles, and meaningfully interact with human participants across different phases and contexts of the collaborative learning journey.

## 3. Literature Review

### 3.1 Collaborative Tools: From Passive Infrastructure to Limited Scaffolding

Early developments in online collaboration tools—such as asynchronous forums, shared documents, and project management platforms—enabled basic resource sharing and communication, but offered minimal intelligent support for the core collaborative process [1]. These tools functioned primarily as passive containers for interaction, facilitating message exchange and file access without actively shaping group dynamics or cognitive engagement [2,3]. However, existing studies largely overlook the extent to which these tools fail to scaffold the social construction of knowledge. While they provide a space for collaboration, they do not dynamically mediate conflicts, suggest integrative perspectives, or adapt to the evolving cognitive needs of learners [4]. A critical limitation lies in their inability to move beyond transactional support to become cognitive partners in the collaborative process. This gap highlights the need for tools that can actively participate in knowledge negotiation—not merely host it.

### 3.2 AI in Education: From Individual Support to Collective Scaffolding

Recent educational AI research has increasingly focused on personalized learning pathways, automated question-answering, and content generation, achieving notable success in supporting individual learners [5,6]. However, these applications largely operate in isolation from the social dynamics of group learning. Few studies have explored how AI can reshape the structure of peer interaction, influence the emergence of collective intelligence, or scaffold the negotiation of meaning across multiple agents [7]. However, existing studies largely overlook the potential of AI to serve as a cognitive partner in collaborative settings. While AI excels at individual adaptation, it remains largely absent from the co-construction of group knowledge. A critical limitation lies in the underdevelopment of hybrid AI systems that integrate technical sophistication with pedagogically meaningfulness. Without embedding AI into the social fabric of learning, we miss opportunities to leverage its power for collective sense-making and bigger conceptual change.

### 3.3 Learning Analytics: From Behavioral Tracking to Cognitive Insight

Learning analytics has evolved to track behavioral indicators such as login frequency, video completion rates, and forum participation [8,9]. While useful for surface-level engagement metrics, these approaches often fail to capture the deeper cognitive and affective dimensions of collaboration—such as the density and evolution of idea conflicts, the pathways to consensus, or the distribution of cognitive contributions across group members [10]. Nonetheless, the integration of advanced natural language processing (NLP) and social network analysis with cognitive theories such as cognitive dissonance, for the purpose of generating actionable insights, remains an underexplored area in the current literature [11]. Current models remain descriptive rather than explanatory, offering little guidance for intervention or design. A critical limitation lies in the fragmentation of methods and the lack of integration between behavioral tracking and cognitive modeling. Only by embedding theoretical insights, such as cognitive dissonance, into analytical models can learning analytics move beyond description to provide explanatory power and actionable feedback for collaborative learning environments.

### 3.4 Conclusion of Literature Review

By mapping these developments thematically, we reveal a clear trajectory: from isolated tools to integrated ecosystems, from individual support to collective scaffolding, and from behavioral tracking to cognitive insight. The proposed framework seeks to consolidate these strands into a coherent, actionable model for future research and development, one that explicitly addresses the gaps identified in collaborative tools, AI's role in groups, and the cognitive depth of learning analytics.

## 4. Discussion & Synthesis

### 4.1 Framework Integration and Theoretical Advancements

The proposed framework advances AI-supported collaborative learning by bridging the persistent gap between theoretical insight, technological design, and analytical depth. It moves beyond prior approaches that treat AI as a passive evaluator or isolated tool, instead positioning it as an intelligent cognitive partner within the learning ecosystem. This integration directly addresses the fragmentation in the literature: it unites the social constructivist vision of knowledge co-construction with the need for authentic intelligent partnership and deep process analytics. Consequently, the framework provides a cohesive new paradigm where AI dynamically scaffolds both the social interactions and cognitive processes fundamental to collaborative learning.

### 4.2 Addressing Identified Gaps Through Core Innovations

To overcome the fragmentation in current systems, this framework reconceptualizes AI-supported learning not as a set of isolated tools, but as a closed-loop cognitive regulatory system. Unlike traditional linear models that treat interaction, analysis, and intervention as separate stages, our framework establishes a continuous feedback loop where each core innovation serves as the input or output of the others, dynamically driving the evolution of collective cognition. First, within the Interaction layer, the multi-agent collaboration architecture acts as the generative engine of cognitive friction. Rather than merely facilitating communication, it is designed to actively surface and sustain productive cognitive dissonance among learners. By creating a dynamic social space where ideas clash and perspectives are negotiated in real-time, this layer transforms passive information exchange into active knowledge construction, thereby establishing the problem or cognitive gap that the system must subsequently address. Second, the Analysis layer, powered by the data-driven analytical engine, provides the perceptual mechanism to understand these emergent problems. This module closes the loop by translating raw interaction data into cognitive insights. Through semantic NLP and network analysis, it quantifies the density of idea conflicts and maps the trajectory of shared understanding. It diagnoses the precise nature and intensity of cognitive dissonance, transforming social behaviors into actionable diagnostic metrics that inform the next stage of system adaptation. Finally, in the Intervention layer, the hybrid LLM orchestration serves as the adaptive resolution mechanism. Acting upon the diagnostic data from the analytics layer, this engine dynamically configures specialized AI models to deliver tailored scaffolding. Whether decomposing complex tasks or mediating conflicting viewpoints, the orchestration system intervenes not with generic prompts, but with strategies specifically calibrated to resolve the identified cognitive gaps. This intervention then feeds back into the Interaction layer, resetting the cognitive state of the learners and triggering the next cycle of collaborative sense-making.

### 4.3 Operational Considerations, Application Scenarios, and Comparative Advantages

However, the framework also presents implementation challenges, including ensuring real-time responsiveness in multi-agent coordination, safeguarding student privacy throughout the analytics pipeline, and fostering pedagogical adoption through teacher training. Future iterations will need to incorporate dimensions such as affective computing and expand to cross-institutional contexts for broader validation. To demonstrate its operability and clarify potential application scenarios, the framework can be implemented in contexts such as a college seminar, where it structures and mediates debate-based learning, or a corporate training program, where it facilitates cross-team knowledge integration. Compared with existing AI-supported learning systems, the proposed framework's primary comparative advantage lies in its holistic integration of multi-agent social dynamics, hybrid AI flexibility, and theory-driven cognitive analytics. It advances the field by providing a coherent design blueprint that systematically connects learning theory, AI technology, and pedagogical practice within a unified ecosystem.

## 5. Conclusion

This study develops a theoretically grounded framework for hybrid large language model–empowered collaborative learning, integrating constructivist learning theory and cognitive dissonance theory. Anchored in a tripartite architecture of multi-agent interaction, hybrid model orchestration, and deep learning analytics, it bridges gaps in intelligent educational systems by enabling deep, dynamic, and equitable collaboration among students, teachers, and AI agents, while generating actionable insights into knowledge co-construction. Theoretically, it reconceptualizes AI-supported collaborative learning as a closed-loop cognitive system, moving beyond the view of AI as a mere tool. It positions AI as an active participant in the learning ecosystem, where interaction, cognitive conflict, and adaptive intervention are co-regulated through human–AI collaboration. Practically, it provides a cohesive model integrating learning theory, AI design, and educational data science, advancing beyond fragmented, linear, or tool-centric approaches to AI in education. This framework is significant in that it amplifies human agency and fosters deeper, more inclusive, and more meaningful collective intelligence. Future work should explore affective computing for emotional and social cue perception, develop secure and interoperable infrastructures, and validate the framework across diverse contexts. Limitations include the need for longitudinal and multi-context empirical studies to establish scalability and generalizability, as well as further refinement of alignment with curriculum goals through teacher training and pedagogical integration.

## References

[1] Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.

[2] Festinger, L. (1957). A theory of cognitive dissonance. Stanford University Press.

[3] Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (pp. 409–426). Cambridge University Press.

[4] Baker, M. J. (2015). Collaboration in collaborative learning. International Journal of Computer-Supported Collaborative Learning, 10(4), 309–348. https://doi.org/10.1007/s11412-015-9222-2

[5] Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), Computer-supported collaborative learning (pp. 69–97). Springer. https://doi.org/10.1007/978-3-642-85098-6_5

[6] Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), Collaborative learning: Cognitive and computational approaches (pp. 1–19). Elsevier.

[7] Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), The Cambridge handbook of the learning sciences (2nd ed., pp. 397–417). Cambridge University Press.

[8] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). On the opportunities and risks of foundation models. arXiv. https://arxiv.org/abs/2108.07258

[9] Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

[10] Ouyang, F., & Jiao, P. (2021). Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence, 2, 100020. https://doi.org/10.1016/j.caeai.2021.100020

[11] Chen, B., & Zhang, J. (2023). Embracing the power of AI in education: A new era of learning. Educational Technology Research and Development. Advance online publication. https://doi.org/10.1007/s11423-023-10285-2

## Funding

This research received no external funding.

## Conflicts of Interest

The authors declare no conflict of interest.

## Acknowledgment

This paper is an output of the science project.