Journal of Education and Educational Research
ISSN: 2957-9465 | Vol. 21, No. 3, 2026

Human–AI Collaboration in Financial English Classroom Reform: An Action Research Study at an Application-Oriented University

Xiaoli Li

School of Pass College of Chongqing Technology and Business University, Chongqing, 401520, China

Abstract: Financial English teaching has long been beset by three persistent problems: students tend to understand terminology only superficially, their genre awareness remains weak, and they struggle to interpret data. Generative AI can offer cognitive scaffolding, but it also brings the risk of cognitive offloading, whereby students let the efficiency of the tool take the place of genuine cognitive engagement. Taking the *Basic Financial English* course in the Business English programme at an application-oriented university as a case, this study draws on distributed cognition and the notion of dual-teacher collaboration. Following the action research spiral proposed by Kemmis, it conducted three rounds of classroom interventions centred on human–AI collaboration. The first round employed a terminology cognitive conflict task to reposition AI; the second adopted reverse deconstruction of genre to establish boundaries in the human–AI division of labour; the third used correction of data narratives to develop critical human–AI collaboration literacy. Data from classroom observation, student reflective journals, AI dialogue logs and student work indicate that human–AI collaboration lowered the cognitive threshold of Financial English, but the quality of collaboration depended on whether the teacher could design structural tension between the efficiency that AI supplies and the cognitive retention that students achieve. The study identifies a progressive mechanism in human–AI collaboration, moving from cognitive offloading through cognitive conflict to cognitive retention, and discusses the conditions and limits under which it can be transferred to Financial English classrooms in application-oriented universities.

Keywords: Human-AI Collaboration; Financial English; Action Research; Cognitive Offloading; Business English.

1. Introduction

Financial English is one of the areas in the Business English curriculum where language ability and disciplinary knowledge are most tightly intertwined. It differs from general English, which centers on language form, and from finance courses taught entirely in Chinese, which aim at conceptual understanding. Students have to handle three cognitive loads at the same time: the precise meaning of financial terms in professional contexts, the information-compressed structure typical of financial discourse, and the narrative logic and value judgement hidden behind data. In application-oriented universities, this difficulty is compounded. Students enter with uneven English proficiency and almost no prior knowledge of finance, while contact hours are limited. The traditional route of "glossary + text translation + exercises" often reaches only the dictionary meaning of terms and rarely develops the ability to use them in context.

The spread of generative AI has introduced a new variable. AI can explain terms instantly, generate sample texts, and help interpret data, so it seems a natural cognitive scaffold for Financial English. Recent ESP research similarly reports that AI-enhanced modules can support financial literacy development among learners (Barseghyan & Hovakimyan, 2026) [1]. Yet preliminary classroom observation raised a warning. When AI is "good enough," students tend to outsource the whole task of understanding: they ask AI to explain terms, summarize paragraphs, and produce analysis. Learning is reduced to a smooth sequence of "input prompt—receive answer," while substantive growth in disciplinary cognition does not necessarily follow. Cognitive science calls this "cognitive offloading": transferring cognitive burden to an external tool. Efficiency rises in the short term, but knowledge retention and deep processing may weaken.

This is the central tension the study tries to address. If human–AI collaborative classroom reform takes "AI is useful" as its only criterion, it may bypass the very cognitive difficulty that Financial English teaching most needs to tackle. The question is not simply whether AI can help students learn Financial English, but under what conditions human–AI collaboration promotes substantive disciplinary cognition rather than dissolving cognitive engagement in the name of efficiency. Taking Basic Financial English as the site, this study designs and examines, through three rounds of action research, a form of "human–AI collaboration with tension."

2. Theoretical Framework and Literature

2.1. From "Tool Mediation" to "Distributed Cognition": Locating Human-AI Collaboration

A common theoretical assumption when bringing AI into teaching is to treat it as tool mediation: teachers use tools, students use tools, and tools serve existing teaching objectives. The limitation of this framework is that it assumes cognition still takes place "inside the head," with tools merely accelerating or facilitating external means. Distributed cognition offers a different view: cognitive activity is not confined to the individual but distributed across people, tools, environments, and social relations. Tools are not passive instruments; they are attentive elements in cognitive processing. When a student works with AI to interpret a financial report, the understanding is not "the student's understanding" plus "AI assistance"; it is the product of a

cognitive system formed by human and non-human actors. Kong (2026) [2] makes a comparable point in proposing a human–AI and interpersonal collaborative foreign language teaching model, in which cognitive work is distributed rather than simply assisted.

This shift has direct methodological implications for Financial English teaching. If cognition is distributed, the key design question is not whether to use AI, but how cognitive tasks are allocated between humans and AI, how boundaries are set, and how one side is prevented from replacing the other's cognitive work. Distributed cognition does not necessarily produce deep learning. Distribution can be productive or undermining. When AI takes over cognitive operations that students should perform, collaboration degenerates into offloading.

2.2. The Cognitive Specificity of Financial English as ESP

Financial English is a cognitively demanding branch of English for Specific Purposes. Its teaching difficulty cannot be reduced to "hard vocabulary" or "long sentences." Existing research points to the compression of information density as a core challenge: annual reports, financial statements, and financial commentary use complex noun phrases, embedded prepositional phrases, and data-embedded expressions to pack multiple layers of information into limited syntactic space. Readers must "unpack" this information and reconstruct logical relations. At the same time, financial terms are highly register-dependent. A word common in everyday English, such as "leverage," "exposure," or "margin," acquires technical meanings in financial contexts, and its use varies subtly across subdomains such as corporate finance, financial markets, and macroeconomics.

This means the cognitive task of Financial English is not a linear "learn words first, then read texts." Students need to move back and forth between terminology, discourse structure, and disciplinary concepts. Traditional teaching slips into glossary memorisation precisely because it lacks sufficient contextual scaffolding for this movement. AI may provide such scaffolding, but it may also let students bypass the movement itself and obtain the conclusion of "what this passage is saying."

2.3. The Risk of "Offloading" in Human-AI Collaboration and the Principle of "Dual-Teacher" Division of Labour

Recent work on human–AI collaboration in foreign language education has begun to address this tension. Zhang's (2026) [3] "dual-teacher, three-language" model in Accounting and Financial English is directly relevant the human teacher takes responsibility for value guidance, complex case analysis, and critical thinking, while the AI teacher handles repetitive Q&A, homework feedback, and personalised practice. The logic behind this division is not a crude split between "AI does the simple things, humans do the complex things." It allocates tasks according to levels of cognitive processing: judgement, weighing, questioning, and meaning construction remain with the human cognitive system, while information retrieval, form checking, and immediate feedback can be assigned to AI. This division is consistent with recent action research on AI-collaborative multi-interaction teaching models, which also stresses task-level rather than tool-level integration (Yang, Zhang, Sun 2025) [4].

Yet the dual-teacher framework still faces an implementation problem. In real classrooms, students do not automatically know when to call on AI, when to switch it off, and when to question its output. The division of labour has to be experienced and internalised in concrete tasks. Action research suits this study because it allows the researcher to design task structures repeatedly in a real classroom, observe students' collaborative behaviour, and adjust the human–AI boundary in response to feedback, rather than presupposing a "correct" model of collaboration from the outside.

3. Research Design

3.1. Method and Site

The study adopts an action research approach, following Kemmis and McTaggart's spiral model: each round includes planning, acting, observing, and reflecting, and each round builds on the reflection of the previous one. The site is the Basic Financial English classroom taught by the author in the School of Foreign Languages of an application-oriented university. The participants are second-year Business English majors. The course is a core course, with two contact hours per week. The action research covers three rounds of intervention over one semester, each lasting about three weeks and embedded in regular teaching. The researcher is also the course teacher. This dual teacher–researcher identity is inherent to action research, and possible bias from the researcher's position was balanced through triangulation of multiple data sources and regular discussions with a critical friend, a colleague from the same teaching office.

3.2. Problem Focus and Task Design across Three Rounds

The three rounds were not fully planned in advance but focused progressively on the basis of each round's reflection. The starting point of Round 1 was the "illusion of AI fluency" observed in earlier teaching: after students used AI to look up terms and generate definitions, they still could not judge in class whether a term fitted a given context. Round 1 therefore focused on the depth of terminology cognition and designed a "terminology cognitive conflict task." Round 2, building on the finding that the smoothness of AI definitions removed cognitive conflict, shifted to genre awareness at the discourse level and designed a "genre reverse deconstruction task." Round 3, after students had gained some experience of human–AI collaboration, moved up to critical judgement in data interpretation and designed a "data narrative correction task." The three tasks shared a progressive goal: from "letting AI help understanding" to "understanding through tension with AI."

4. Process and Findings of the Action Research

4.1. Round 1: Terminology Cognitive Conflict—Making AI's "Fluency" Questionable

The core problem of Round 1 was that instant AI definitions made terminology learning too smooth, so students skipped the cognitive step of testing meaning in context. The task was designed as "terminology cognitive conflict." Three terms whose meanings shift across financial subcontexts were selected from the textbook (for

example, margin in retail and in financial derivatives). Students first wrote their own understanding based on the context, then asked AI for an explanation, and then completed a conflict log: they identified inconsistencies between AI's explanation and their own understanding, or between AI's explanations in different contexts, and explained in groups which explanation better fitted the current text and on what grounds. AI's role was deliberately limited to "a questionable source of definitions" rather than "the provider of correct answers."

Action and observation. Classroom observation showed the expected differentiation. About one-third of the students recognized AI's context sensitivity and wrote in the conflict log: "AI says margin is profit, but the text is about derivatives margin; AI's first explanation does not apply." More students, however, displayed another pattern. Once AI gave a clear and well-organized explanation, they abandoned their own vague understanding and adopted AI's version directly. The conflict log became a formality ("AI's explanation is more accurate"). In reflective journals, one student wrote: "I had guessed wrong at first, but AI explained it clearly, so there was no conflict." This comment reveals the problem: AI's smooth output did not create cognitive conflict; it removed conflict. Students experienced "my understanding was corrected by AI" as learning having taken place, while the process of correction was not internalized.

Round 1 shifted the research focus from whether AI provides correct explanations to whether the way AI intervenes preserves cognitive friction. If AI always appears as the authoritative explainer in terminology tasks, collaboration will inevitably slide into offloading. The decisive adjustment for Round 2 was to change AI's position in the task structure, from "providing explanations" to "exposing structure." Students would no longer face AI's answers but the way AI-generated texts are constructed, so that cognitive work moved from "accepting explanations" to "analysing how explanations are organized."

4.2. Round 2: Genre Reverse Deconstruction—Seeing Structure in AI Output

The genre features of financial discourse—information compression, embedded data, and the interweaving of evaluative and factual language—are an important source of reading difficulty, but genre analysis in traditional teaching is often demonstrated by the teacher and passively received by students. Round 2 designed a "genre reverse deconstruction task." For a real excerpt of financial text (for example, an excerpt from a listed company's Management Discussion and Analysis), AI first generated a "simplified" or "summarized" version. Students then completed two comparative analyses: (1) which information in the original was combined, omitted, or reordered in the AI version; and (2) how AI's simplification changed the original's argumentative logic or risk signals. The student output was not a "summary" but an analytical account of what AI did, what it did not do, and what consequences its choices produced.

This task structure produced visibly different collaborative behaviour from Round 1. Because AI moved from "answer provider" to "object of analysis," students had to enter the text: to judge what AI had omitted, they first had to understand how the original organized information. Classroom discussion produced types of utterance absent in Round 1: "AI split this long sentence into three, but the original used nomination to compress causality together. Once it is split, 'because—therefore' becomes too explicit; the original was hinting rather than asserting." This suggests that students began to shift attention from "what the meaning is" to "how the meaning is constructed." Analysis of AI dialogue logs confirmed the change. In Round 1, many prompts were "what does this term mean"; in Round 2, more prompts became "why did you omit this sentence," "what did you translate this phrase into, and are there other possibilities?" Students began to ask about AI's operations themselves.

The success of Round 2 lay in the fact that genre deconstruction provided an operable middle-level structure for the human–AI division of labour: AI was responsible for "simplifying," and students for "examining the cost of simplification." But this division still centred on analysis within the text and had not yet touched a more challenging dimension of Financial English: data. Numbers in financial texts are never neutral. Which number is presented, against what benchmark it is compared, and in what rhetorical frame it is wrapped all carry narrative intent. If students learn only to analyse discourse structure while continuing to outsource data interpretation to AI, human–AI collaboration remains at the language level. Round 3 therefore moved from "textual structure" to "data narrative."

4.3. Round 3: Data Narrative Correction—Forming Judgement through Human–AI Disagreement

Round 3 was designed as "data narrative correction." A set of real financial data was provided (for example, a comparison of revenue growth and profit margins between two companies in an industry). AI first generated an analysis, and students were then asked to find "another story that the same data could tell." The steps were: (1) identify implicit causal assumptions or evaluative leanings in AI's analysis; (2) use the data themselves to test whether the assumption was the only reasonable one; and (3) write an alternative analysis and explain under what conditions their version would be more appropriate than AI's. AI's role in this round was "an analyst to be questioned and talked back to," not a tool or an answer key.

Student reflective journals and work analysis showed that Round 3 touched the most valuable cognitive moment in human–AI collaboration: when AI's output is not "wrong" but "plausible while carrying a particular frame," students must mobilise financial concepts, data logic, and rhetorical awareness to judge. One student recorded: "AI said Company A had higher growth, so it was more worth investing in. But when I put profit margin in, the high growth might simply come from consolidation, while the margin was actually falling. AI did not lie, but it chose a telling that made the story sound better." Other students showed difficulty: when they found AI's analysis "reasonable," they had little motivation to produce an alternative narrative and tended to accept the first coherent story. This split suggests that critical human–AI collaboration is not automatically acquired. It needs the experience of "examining AI's operations" accumulated in the first two rounds.

Looking back, a clear line emerges. Round 1 exposed the problem of "fluency dissolving cognitive conflict." Round 2 rebuilt cognitive friction by turning AI from explainer into object of analysis. Round 3 extended that friction from discourse structure to data narrative, raising human–AI collaboration from "assistance in understanding" to "practice in judgement." What the three rounds jointly point to is not

"optimizing the use of AI functions" but maintaining a productive cognitive tension between humans and AI.

5. Discussion

5.1. "Cognitive Offloading—Cognitive Conflict—Cognitive Retention": A Progressive Mechanism

The three rounds can be distilled into a progressive mechanism. Cognitive offloading is the starting point and the risk: the immediacy and fluency of AI naturally tend to transfer cognitive burden away from learners, and without intervention in task design, offloading becomes the default. Cognitive conflict is the key lever for intervention: only when AI's output is no longer an "answer" but an "object to be examined" are students forced into cognitive operations between AI's supply and their own judgement, and conflict becomes the trigger for deeper processing. Cognitive retention is the criterion for judging the quality of collaboration: retention is not about what students "know," but whether they have acquired analytical frameworks and habits of judgement that can be called upon when AI is not present. This mechanism echoes existing discussions of the risk of cognitive dependence in "human–AI and interpersonal collaboration": preventing offloading does not depend on restricting AI use, but on redesigning AI's position in the task structure, so that it moves from "substitute for cognition" to "interlocutor in cognition."

5.2. The Irreplaceable Function of the Teacher in Human–AI Collaborative Classrooms

The three rounds also reveal something easily obscured by the discourse of "AI empowerment": the quality of human–AI collaborative classrooms depends less on AI's capabilities than on the teacher's ability to design cognitive tension. In each round, the teacher's core work was not "teaching what AI can teach," but doing what AI structurally cannot do: identifying unarticulated areas of confusion in students' cognition , designing task frameworks that make AI's output questionable, and providing counterexamples and follow-up questions when students were inclined to accept AI's coherent narrative. This is consistent with the dual-teacher framework's placement of value guidance, complex case analysis, and critical thinking with the human teacher, but it further shows that this division is not automatically realized by "adding AI" to an existing classroom. It has to be practised repeatedly through the mediation of task structure. In a human–AI collaborative classroom, the teacher's primary professionalism lies not in proficiency in operating AI but in judging at which cognitive point AI should "step back."

5.3. Conditional Implications for Financial English Teaching in Application-Oriented Universities

The practice has reference value for Financial English teaching in Business English programme at application-oriented universities, but its implications are conditional rather than a model to be copied directly. First, the premise of human–AI collaborative reform is that the teacher accurately identifies the "cognitive bottlenecks" in the course. If the teacher is not clear whether the difficulty of Financial English lies in terminology, genre, or data logic, AI's intervention can only be generic "assistance" and cannot form structural tension. Second, the feasibility of the three rounds depended on a relatively small class size and institutional space for the teacher to make within-course adjustments. Under tight contact hours and rigid content requirements, embedded action research needs a more streamlined design. Finally, critical human–AI collaboration literacy develops cumulatively: the judgement difficulties in Round 3 suggest that if students have not had experience of "examining AI output" in earlier Financial English learning, it is unrealistic to expect cognitive upgrading within one round of tasks.

6. Conclusion

This study took the Basic Financial English classroom as its site and carried out three rounds of action research, starting with terminology cognitive conflict, moving through genre reverse deconstruction, and ending with data narrative correction, in order to construct and examine a human–AI collaborative classroom organised around cognitive tension. The central finding is not that human–AI collaboration outperforms traditional teaching. Rather, its educational value is conditional: it is productive when it encourages deep cognition and counterproductive when it takes over cognitive processes. The key condition is whether the task structure shifts AI from a source of answers to an object that students examine and talk back to. For Financial English teaching in Business English programme at application-oriented universities, this means reform should not stop at choosing AI tools and training students to use them; it should move on to redesigning cognitive tasks and managing human–AI boundaries dynamically. This conclusion also points to further questions: how to keep cognitive tension workable in larger classes, and how to make the development of critical human–AI collaboration literacy something more than an event in a single action research project, so that it becomes a thread running through the curriculum.

References

[1] Barseghyan, L., & Hovakimyan, L. (2026). AI-Enhanced Module for ESP Learners: The Case of Financial Literacy. Foreign Languages in Higher Education, 30(1), 57–80. (in Chinese)

[2] Kong, L. (2026). Constructing a Human–AI and Interpersonal Collaborative Foreign Language Teaching Model. Foreign Language Teaching and Research, 58(1), 108–120. (in Chinese)

[3] Zhang, X. (2026). Teaching Concepts and Instructional Design of Accounting and Financial English for Competence Development. Lecture presented at the School of Accounting, Lanzhou Technology and Business University, Lanzhou, China.

[4] Yang, X., Zhao, Z. J., & Sun, Q. Y. (2025). Action Research on College English (Integrated Course) under an AI-Collaborative Multiple-Interaction Teaching Model. Overseas English, (12), 11–13. (in Chinese)

[5] Luo, X. H., & Liu, L. L. (2026). Construction and Empirical Study of Business English Course Ecosystem in Application-Oriented Universities Driven by Industry-Education Integration and GenAI. Creative Education Studies, 14(3), 200–207.

[6] Zhao, J. J. (2025). Financial English in the Digital Age. China Financial and Economic Publishing House. (in Chinese)

[7] Sesriyani, L. (2026). Integrating Artificial Intelligence (AI) in ESP Instruction: Enhancing Digital Business English Literacy Among Economic Education Students. Global Expert Journal of English.

[8] Kemmis, S., & McTaggart, R. (Eds.). (1988). The action research planner. Deakin University Press.