Enhancing vocational students' professional competencies through AI agent-supported human–AI collaborative learning: A longitudinal mixed-methods study

Tan Huizhu, (ORCID: 0009-0001-4769-2265), Shenzhen Polytechnic University, China

Abstract

While generative AI (GenAI) integration in education often remains at the tool-assistance level, this study proposes and evaluates an AI agent-based collaborative teaching system designed to operationalize structured human-AI interaction in vocational training. Utilizing a multi-agent architecture featuring role-specialized agents—customer simulation, decision support, and instructional guidance—the system implements a triadic framework that enables coordinated engagement among students, instructors, and AI agents across the entire learning cycle. A mixed-method longitudinal evaluation of a "Cross-Border E-Commerce Customer Management" course (2022–2025, N=232) revealed a significant upward trend in instructional quality, with average teaching evaluation scores rising from 90.58 to 93.81. These results demonstrate that agent-oriented design and iterative optimization effectively transform GenAI from an auxiliary tool into a collaborative educational partner. Despite current technical limitations in emotional intelligence, the proposed architecture offers a replicable, human-centered framework for interactive competency development and instructional optimization in vocational education and related skill-based domains.

Keywords: Generative artificial intelligence, AI agents, human–AI collaboration, vocational education, instructional quality, cross-border e-commerce

1 Introduction

The rapid expansion of digital commerce has increased the demand for professionals capable of operating in complex and continuously evolving online business environments. Vocational education is therefore expected to move beyond knowledge transmission toward competency-oriented training that emphasizes authentic problem solving and practical decision-making. However, providing effective experiential learning opportunities remains challenging in classroom settings, where instructional time and realistic practice environments are often limited. Experiential learning theory suggests that professional competence develops through iterative cycles of action, reflection, and adaptation, highlighting the importance of learning through experience rather than passive instruction (Kolb & Kolb, 2005). Prior research in artificial intelligence in education has explored intelligent tutoring systems and adaptive learning environments that support personalized guidance and interactive learning processes

(Roll & Wylie, 2016; Holmes et al., 2019). Recent studies further examine the role of large language models in facilitating human-AI collaborative learning, highlighting both opportunities and pedagogical challenges associated with generative AI integration (Kasneci et al., 2023). These studies indicate that the effectiveness of AI in education depends not only on technological capability but also on instructional design and interaction structure (Kasneci et al., 2023; Roll & Wylie, 2016).

In vocational domains such as cross-border e-commerce, learners must repeatedly analyze contextual information, evaluate operational strategies, and respond to dynamic scenarios. Yet traditional instructional approaches, including case-based teaching and static simulations, frequently fail to provide sustained interaction and continuous feedback. As a result, learning activities may remain episodic, limiting opportunities for learners to refine decision-making strategies through repeated practice. These challenges have motivated increasing interest in technology-enhanced learning environments capable of supporting authentic and interactive learning experiences.

Recent advances in generative artificial intelligence (AI) have introduced new possibilities for supporting teaching and learning processes. Studies have shown that generative AI systems can assist learners through feedback generation, content support, and interactive dialogue, thereby reshaping how students engage with knowledge and learning tasks (Yan et al., 2023). At the same time, international policy discussions emphasize that AI should augment rather than replace human learning processes, encouraging human-centered approaches that preserve learner agency and instructional guidance (UNESCO, 2023). Emerging research further highlights the growing importance of human-AI collaboration, showing that students increasingly interact with AI as a learning partner rather than merely a tool (Razmerita, 2024).

Despite these developments, many existing educational applications employ generative AI primarily as a general-purpose assistant. Such implementations often lack structured pedagogical design and clearly defined instructional roles, which may lead to fragmented learning experiences and unclear interaction patterns. Recent studies on human-AI collaborative learning suggest that structured interaction design and role differentiation may be critical for supporting higher-order learning and sustained engagement, particularly in vocational education contexts (Hwang & Lee, 2025).

To address this gap, this study investigates how structured human-AI collaboration can enhance experiential vocational learning through the following research questions.

1. How does the proposed AI agent-based collaborative teaching system affect students' perceived instructional quality over time in a vocational training context?
2. What are students' usage patterns and perceptions of human-AI collaboration when interacting with role-specialized AI agents (customer simulation, decision support, instructional guidance)?
3. What are the perceived strengths and limitations of the triadic (student-instructor-AI) collaborative model, as reflected in students' qualitative feedback?

To answer these questions, an agent-supported collaborative learning environment was designed in which multiple AI agents assume specialized instructional roles to facilitate simulation-based activities, guided practice, and continuous feedback. Rather than replacing instructors, the proposed approach redistributes instructional responsibilities between human teachers and AI agents, enabling sustained interaction while maintaining pedagogical oversight.

This study contributes to research in educational technology and AI-supported learning in three primary aspects. First, it proposes a human-AI collaborative learning model operationalized through role-specialized AI agents within vocational training contexts. Second, it examines how agent-mediated interaction reshapes learning processes by enabling sustained experiential engagement and scaffolded decision-making. Third, it provides empirical evidence from an authentic cross-border e-commerce training course, demonstrating improvements in perceived instructional quality and learner perceptions toward human-AI collaboration.

2 Literature Review and Theoretical Framework

With the explosive development of generative AI technologies, exemplified by large language models (LLMs), the education sector is undergoing a profound paradigm shift. In its early stages, AI applications in education primarily focused on tool-empowering functions such as personalized recommendations and intelligent grading (Roll & Wylie, 2016). However, marked by the emergence of Large Language Models (LLMs), generative AI is driving a shift in its role from an auxiliary tool to a quasi-autonomous participant in teaching activities (Kasneci et al., 2023). The academic community widely recognizes that the subject structure of teaching activities is evolving from the traditional "teacher-student" dual structure to a "teacher-student-AI" triadic structure (Yi, 2025).

Regarding the systemic elements of this triadic teaching structure under AI, Yikai Yu et al. proposed an eight-element model for human-AI collaborative teaching grounded in humanistic principles. This model integrates AI agents as a new subject element within the teaching system, emphasizing the dynamically weighted collaborative relationship among teachers, students, and AI. It provides the core framework for constructing the "teacher-student-AI" triadic interaction mechanism in this study (Yi, 2025). Lu Yu et al. constructed an evolutionary pathway progressing from labor substitution to human–AI collaboration, ultimately advancing toward cognitive integration (Lu & Tang, 2025).

Synthesizing the above strands, three recurring themes emerge: (a) AI's role is shifting from a passive tool to an active pedagogical agent, (b) effective integration requires structured interaction designs rather than generic AI assistance, and (c) vocational domains particularly benefit from high-fidelity, low-risk simulation environments. However, empirical evidence on how role-specialized AI agents support sustained, competency-oriented learning in vocational settings remains scarce. Most existing studies focus on general higher education or short-term interventions, leaving a gap in longitudinal, mixed-methods investigations of multi-agent collaborative systems. This study directly addresses that gap.

Current research frontiers focus on exploring "human-machine symbiosis." This transcends master-servant or competitive dynamics, emphasizing a partnership where humans and AI—as heterogeneous agents with distinct strengths and limitations—interdependently co-evolve toward shared educational goals (Wang & Deng, 2025). Within this relationship, teachers must transition from knowledge transmitters to reflective facilitators within human-AI collaborative education, while students must strengthen their agency as self-directed learners.

The emergence of generative AI offers revolutionary tools to address various pain points in cross-border e-commerce training. Its enabling opportunities primarily manifest in: 1) Scenario Simulation: AI can "dynamically update knowledge bases to precisely replicate real-world transaction scenarios," cost-effectively constructing diverse training environments covering customer communication, logistics disputes, and social media PR.2) Skill Augmentation: AI serves as an "AI Skills Coach," assisting students in market trend analysis, product copywriting, and customer insight development to enhance professional competencies and data sensitivity. 3) Instructional Reconstruction: Institutions are pioneering "AI+Trade" interdisciplinary courses, leveraging AI to redesign digital teaching environments and assessment systems, thereby advancing industry-education integration.

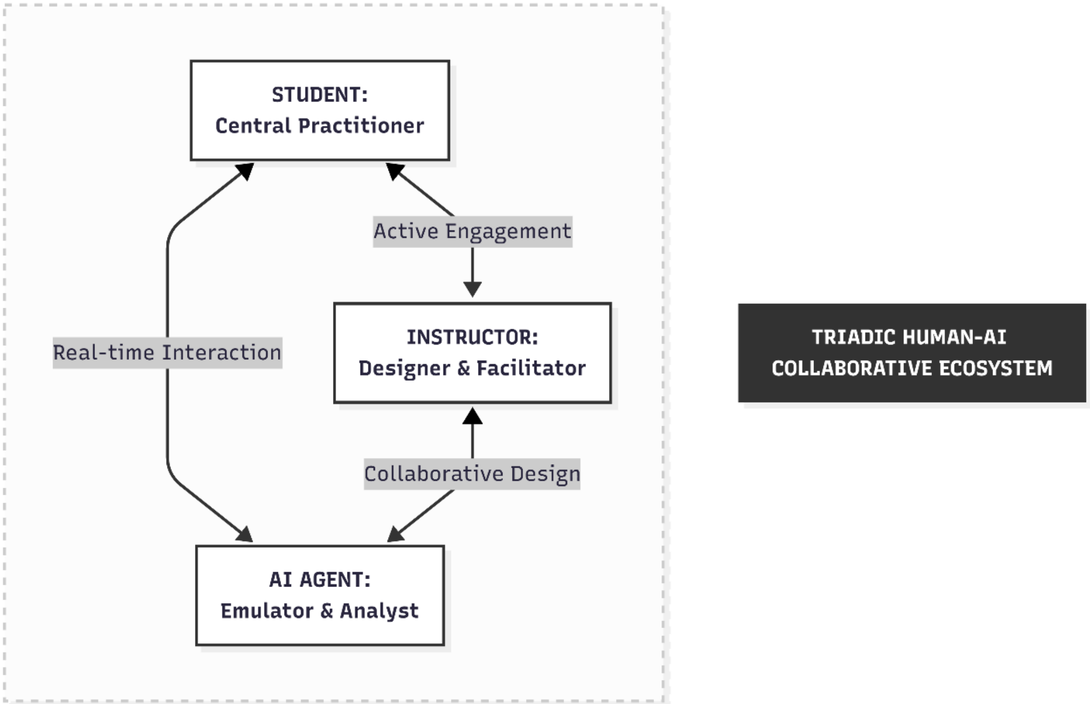
This study takes the Cross-Border E-Commerce Customer Management course in Shenzhen Polytechnic University as a case example. Based on the above "human-machine symbiosis" collaborative teaching philosophy, it constructs a "Teacher-Student-AI" triadic human-AI collaborative teaching model. During training, AI agents can assume multiple roles: acting as "AI customers" or "AI suppliers" to create simulated dialogue environments, serving as "AI data analysts" to provide decision support, and functioning as "AI business mentors" to offer procedural guidance. Teachers transition from facilitators to designers of learning scenarios, coordinators of human-machine interactions, and enablers of higher-order thinking and emotional value. To ensure the effectiveness of AI simulations, addressing their "mechanical feel" is crucial. By designing anthropomorphic social cues—such as character backgrounds, linguistic styles, and emotional expressions—we enhance the AI's "social presence," enabling students to perceive it more naturally as an interactive partner and thereby stimulating deeper learning engagement. Teachers focus on domains where AI is less proficient: guiding students in critical thinking (evaluating the validity of AI suggestions), value judgments (addressing ethical and cultural conflicts), strategic innovation (designing solutions beyond templates), and providing emotional support and humanistic care. This achieves "innovative empowerment for higher-order tasks" while upholding the core mission of education.

3 The AI Agent-Based Collaborative System: Architecture and Design

3.1. System architecture and agent-oriented design

The core objective of the "Agent Simulation + Teacher Collaboration" teaching model developed in this study is to break away from the simplistic "student-software (or case)" binary interaction in traditional practical training. It aims to establish a student-centered, "triadic human-AI collaborative" learning ecosystem where both teachers and AI agents provide dynamic support.

Figure 1. The triadic human–AI collaborative (HAC) framework



The model operates according to four core design principles:

1. Student-Centered Principle: All instructional design revolves around the ultimate goal of enhancing student competencies. The configuration of AI agents and teacher roles creates a "low-risk, high-fidelity, strong-feedback" practice environment to stimulate students' proactive exploration and problem-solving abilities.

2. Human-AI Complementarity Principle: Clearly delineates the respective domains of AI and teachers. AI excels at providing unlimited, standardized, and instantaneous simulated interactions and information processing; teachers lead in value judgment, strategy refinement, emotional support, and innovation inspiration. The two are not substitutes but achieve synergistic teaching enhancement where "1+1>2" through collaboration.

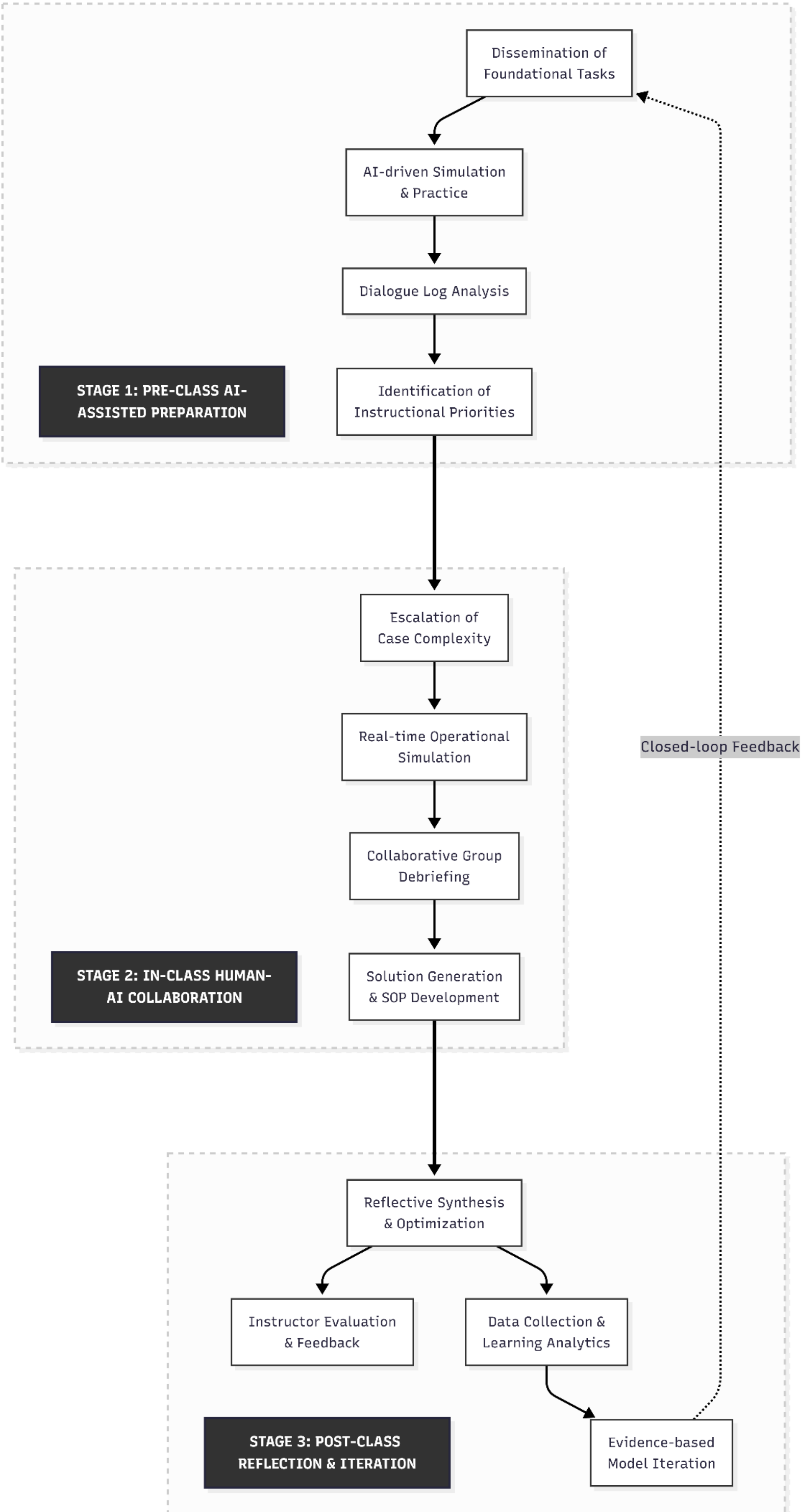
3. Closed-Loop Task Scenario Principle: Teaching tasks simulate the complete closed-loop of real cross-border e-commerce customer management—from demand insight and communication to solution generation and post-action review. AI agents are integrated throughout, ensuring students receive support at every stage and forming a

coherent chain of competency training.

4. Data-Driven Iteration Principle: The model incorporates self-optimization mechanisms. It continuously collects process data (e.g., dialogue rounds, assistance hotspots, solution adoption rates) and outcome feedback (e.g., satisfaction scores, competency self-assessments) generated during student-AI interactions. This data dynamically adjusts the AI agent's response logic, enriches scenario libraries, and guides instructors in refining teaching priorities.

The following example illustrates how the three-stage teaching process operates in a real classroom setting, using a complete cross-border logistics dispute resolution training unit.

Figure 2. The three-stage instructional lifecycle of AI agent integration



Dissemination of Foundational Tasks
AI-driven Simulation & Practice
Dialogue Log Analysis
Identification of Instructional Priorities
STAGE 1: PRE-CLASS AI-ASSISTED PREPARATION
Escalation of Case Complexity
Real-time Operational Simulation
Collaborative Group Debriefing
Solution Generation & SOP Development
STAGE 2: IN-CLASS HUMAN-AI COLLABORATION
Closed-loop Feedback
Reflective Synthesis & Optimization
Instructor Evaluation & Feedback
Data Collection & Learning Analytics
Evidence-based Model Iteration
STAGE 3: POST-CLASS REFLECTION & ITERATION

The first stage is the pre-class AI-assisted preview session. Here, students receive a foundational case task on the training platform—such as an inquiry email from a U.S. customer regarding a one-week package delay. Their task involves practicing initial email responses with an AI-simulated customer. Throughout this process, the AI agent serves as an indefatigable practice partner, providing an interactive dialogue environment. Teachers review all students' initial dialogue records through the backend system, quickly identifying common pitfalls across the class—such as omitting time differences explanations or failing to proactively provide tracking links. These precisely identified weaknesses become the key basis for designing in-class teaching priorities.

The second phase involves human-machine collaborative training during class—the core of the model. Teachers first escalate the scenario by introducing complexity: announcing that the customer posted negative reviews on social media due to unsatisfactory responses, escalating the conflict into a real-time heated exchange with online customer service. Teachers briefly explain the typical nature and core handling principles of logistics crises in cross-border e-commerce, establishing an action framework for students. This is followed by a real-time simulation exercise where students work in pairs: one acts as the customer service representative engaging in high-fidelity text-based dialogue with an AI-simulated angry customer, while the other serves as a supervisor using AI data decision tools to rapidly access the customer's order history and value, providing internal data support to the frontline representative. This segment authentically replicates the pressure and collaboration found in real-world work scenarios.

Following the simulation, the class transitions to a strategy discussion and debriefing phase. Each team presents their dialogue records and resolution strategies, with the instructor guiding a deep-dive class discussion. Key discussion points include: which specific phrases proved effective in calming customer emotions; which emergency commitments carry potential fulfillment risks; where the AI-simulated customer behavior deviates from reality; and how to adjust communication strategies to address these limitations. Through comparative analysis and critical reflection, students distilled fragmented, intuitive experiences into transferable rational strategies. The concluding phase involved summarization and solution generation, where instructors synthesized and elevated best practices. Students then utilized low-code solution generation tools to codify the critical steps, core scripts, and internal collaboration processes from this crisis response into a draft Standard Operating Procedure (SOP) for logistics delay crisis management—a clear, structured document. This marked a pivotal leap from practical application to methodological framework.

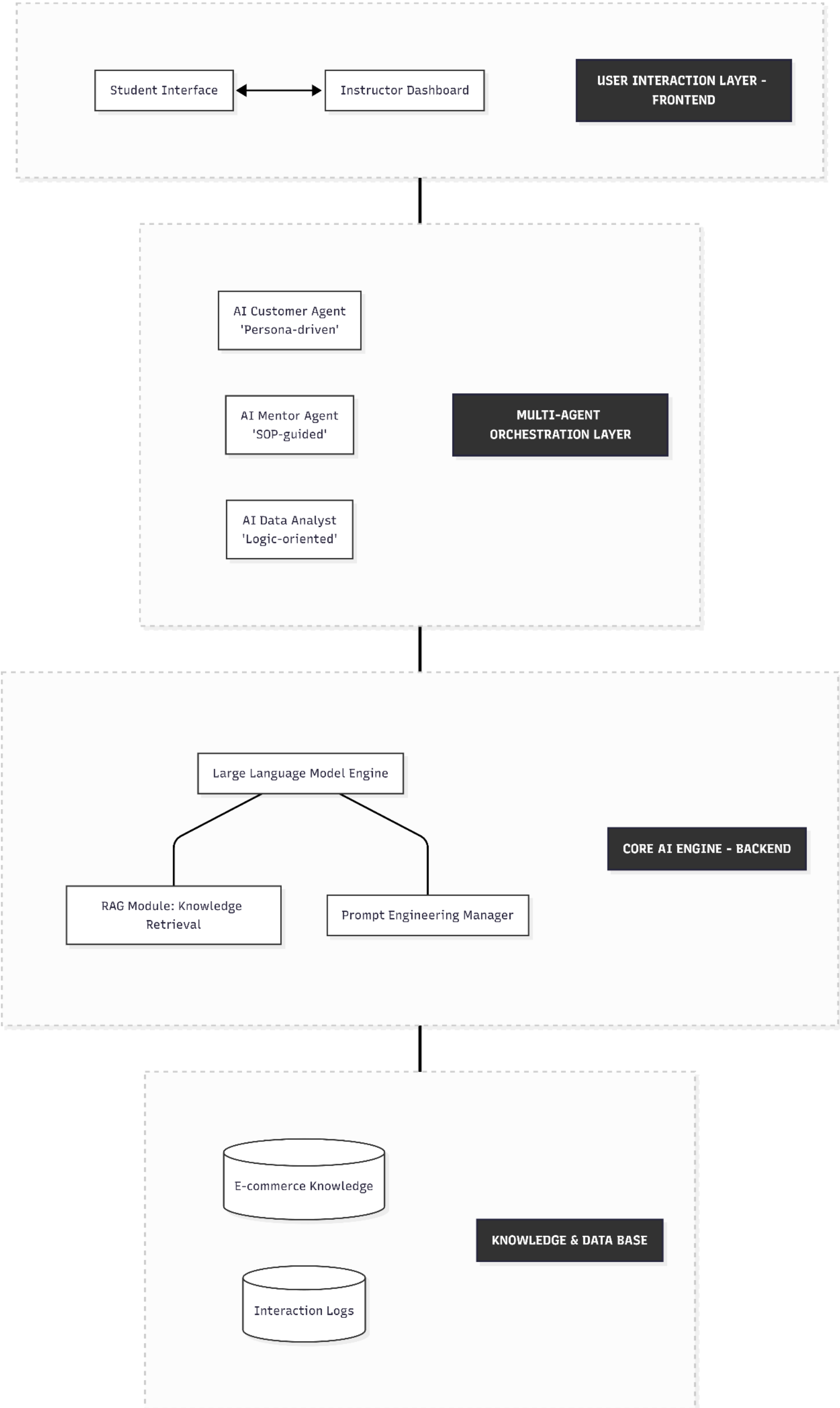
The third phase involves post-class reflection and extension. Students refine their SOP based on classroom feedback and submit a structured reflection report, deeply analyzing their strengths and areas for improvement during the process. Instructors review these reports, providing personalized feedback focused on individual development. Simultaneously, the teaching system automatically collects and analyzes high-frequency questions repeatedly posed to the AI during the training, such as specific compensation proposals or exact delivery timelines. After review by the teaching team, this data serves as direct input for optimizing the AI agent's knowledge base and developing future training scenarios, thereby driving the entire teaching model into an evidence-based closed-loop iteration and continuous optimization cycle.

3.2 Triadic human-AI collaborative pedagogical model

This section focuses exclusively on the system-level technical implementation of the proposed AI Agent-Based Collaborative Teaching System. The technical architecture was derived from pedagogical requirements aimed at supporting experiential learning, scaffolded decision-making, and sustained learner engagement. While the pedagogical workflow and instructional design are described in Section II, the present section details the architecture, agent formalization, knowledge integration mechanisms, and interaction interfaces that enable the operationalization of human-AI collaboration. The design of the multi-agent architecture in this study aligns with the vision of Hwang et al. (2020) who emphasized that the true potential of AI in education lies in its versatile roles as an intelligent tutor, tutee, or learning companion. Our system operationalizes these roles through specialized agents that facilitate high-fidelity simulation training.

To ensure the high-fidelity simulation of cross-border e-commerce scenarios, the system is built upon a Multi-Agent Architecture powered by Large Language Models (LLMs). The technical implementation focuses on transforming the generative capabilities of LLMs into specialized educational subjects through a structured Prompt Engineering framework.

Figure 3. Multi-layered technical architecture of the AI agent-based teaching system


Student Interface
Instructor Dashboard
USER INTERACTION LAYER - FRONTEND
AI Customer Agent 'Persona-driven'
AI Mentor Agent 'SOP-guided'
AI Data Analyst 'Logic-oriented'
MULTI-AGENT ORCHESTRATION LAYER
Large Language Model Engine
RAG Module: Knowledge Retrieval
Prompt Engineering Manager
CORE AI ENGINE - BACKEND
E-commerce Knowledge
Interaction Logs
KNOWLEDGE & DATA BASE

1. Agent Role-Playing and Persona Engineering

The system configures distinct AI Agents by injecting specialized System Prompts and meta-instructions. 1) AI Customer Agent: Integrated with persona-based parameters including cultural background (e.g., North American or European communication styles), specific emotional states (e.g., angry or inquisitive), and historical transaction data. 2) AI Data Analyst Agent: Designed with logical processing prompts that enable the agent to interpret structured logistics data and customer value metrics, providing real-time decision support to students. 3) AI Mentor Agent: Programmed with a domain-specific knowledge base containing Standard Operating Procedures (SOPs) for cross-border trade and platform-specific policies.

2. Knowledge Integration and Context Management

The system employs a Retrieval-Augmented Generation (RAG) approach to mitigate hallucinations and ensure professional accuracy. The "scenario library" comprises real-world industry cases such as logistics disputes and social media PR crises. To maintain the continuity of complex negotiations, a Short-term Memory Buffer manages dialogue history, allowing the AI to exhibit context-aware responses during high-intensity simulations.

3. Data Logging and Closed-Loop System Optimization

To support continuous system optimization, the platform implements a data-logging and feedback pipeline that captures both interaction-level and outcome-level data. Logged data include dialogue turns, agent invocation frequency, assistance request hotspots, and solution adoption patterns.

More importantly, the collected data serve as inputs for iterative system refinement, including prompt optimization, scenario library expansion, and agent response calibration. This closed-loop mechanism allows the system to evolve beyond static rule-based behavior, gradually aligning agent responses with authentic business practices. As a result, the platform supports not only individual course delivery but also long-term institutional knowledge accumulation.

4 Research Methodology

4.1 Research context and participants

The proposed system was deployed in a mandatory "Cross-Border E-Commerce Customer Management" course. A convenience sampling method was employed, involving all students officially enrolled in this course across four consecutive academic years (2022–2025). This approach ensured that the participants represented the typical demographic and skill level of vocational students in this domain.

To balance data breadth and depth, a mixed-methods research design was adopted, drawing on three core data sources:

1. Longitudinal Teaching Evaluation Data: Official student evaluation of teaching (SET) scores from 2022 to 2025 (N=232) were collected to measure the macro-level impact of the system on instructional quality and student satisfaction.
2. Qualitative Textual Data: 35 valid reflections from the "AI Training Learning Reflections" questionnaire were analyzed to capture subjective perceptions of the human-AI collaboration.
3. Quantitative Survey Data: 117 valid responses from the "AI Agent Simulation Training Feedback" survey were used to analyze specific usage behaviors and competency gains.

4.2 Mixed-methods design

Regarding analytical methods, a collaborative human-AI thematic analysis was applied to the qualitative text data. The process was conducted as follows: First, the researcher and an AI analysis assistant performed repeated readings of all texts to gain familiarity. Meaningful statements were then extracted and subjected to open coding. To ensure interpretive validity and coding consistency, the researcher cross-checked the AI-generated codes against a randomly selected subset of the data, achieving high thematic alignment. Through an iterative process of discussion and refinement, these initial codes were grouped into three core themes: "usage behavior," "experience evaluations," and "improvement suggestions."

Quantitative data were analyzed using SPSSAU, a professional statistical analysis platform. For the longitudinal teaching evaluation data, descriptive statistical analysis and trend analysis were employed to examine the trajectory of instructional quality across the four-year implementation. Given that SET scores are standardized institutional metrics, the consistent upward trend in mean values (from 90.58 to 93.81) was utilized as a primary indicator of pedagogical acceptance and system effectiveness. For the survey data, descriptive statistics (frequency and percentage) were used to reveal prevailing trends and competency gains. The results were integrated through a triangulation approach, ensuring the validity and depth of the findings.

5 Results

5.1 Longitudinal analysis of instructional efficacy

Before analyzing specific usage behaviors and qualitative perceptions, we first examine the macro-level instructional efficacy to assess whether the system's implementation translated into measurable improvements in teaching quality.

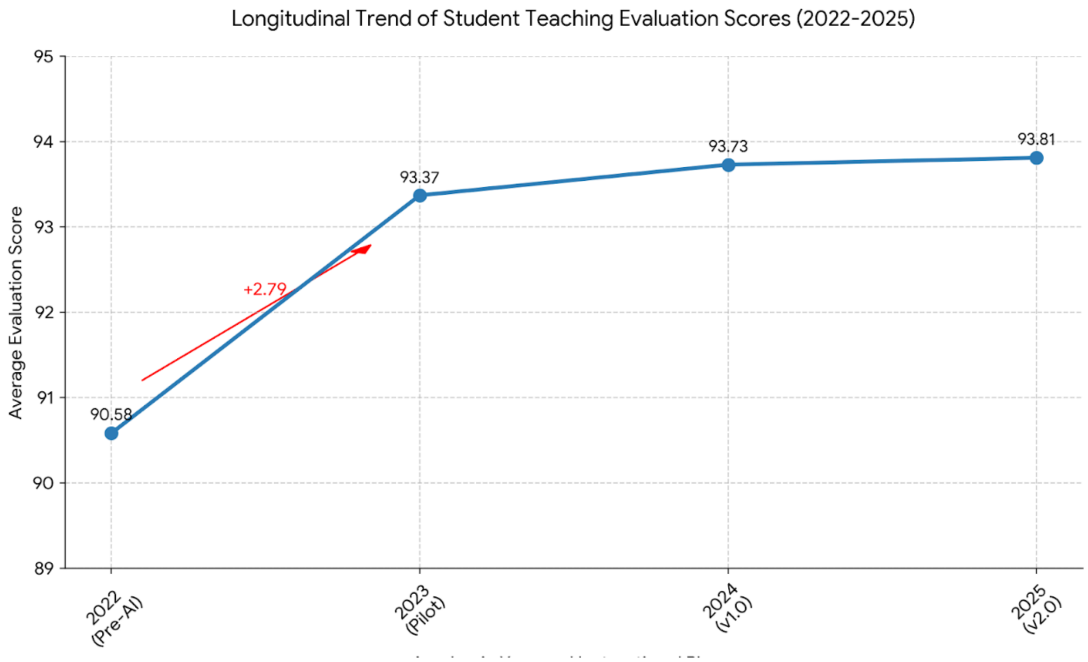
Student teaching evaluation scores were selected as the primary indicator of instructional effectiveness for three reasons. First, unlike task-specific grades, evaluation scores reflect an integrated, learner-perceived measure of overall teaching quality, engagement, and satisfaction—outcomes directly relevant to the human–AI collaborative model. Second, these scores are institutionally standardized and collected longitudinally, enabling consistent comparison across four cohorts. Third, while objective competency tests would be valuable, the course's authentic assessment already involves complex simulated tasks; adding pre-post tests risked disrupting instructional flow. Nevertheless, the study supplements evaluation scores with qualitative reflections and survey data to triangulate findings, and the limitations of this choice are addressed in Section 6.3.

As shown in Table 1, the instructional performance exhibited a steady and significant improvement following the iterative introduction and optimization of AI Agents.

Table 1. Longitudinal Comparison of Student Teaching Evaluations (2022–2025)

Academic Year	**Instructional Phase**	**Sample Size (N)**	**Average Evaluation Score**	**Growth/Shift**
2022	Pre-AI	57	90.58	--
2023	Post-AI	62	93.37	+2.79
2024	Post-AI	63	93.73	+0.36
2025	Post-AI	50	93.81	+0.08

Figure 4. The longitudinal trajectory of student teaching evaluation scores across different implementation phases (2022–2025)



The data reveals two critical trends. First, as visualized in Figure 1, the most significant surge in scores occurred during the pilot phase (between 2022 and 2023, an increase of 2.79 points), coinciding with the initial transition from traditional lecturing to the "AI Simulation + Teacher Collaboration" model. This immediate positive feedback suggests that the high-fidelity scenarios provided by AI agents effectively addressed the lack of interactive practice in traditional vocational training. Second, following this initial surge, the evaluation scores entered a sustained high plateau (above 93.70) as the system matured. This sustained excellence reflects the successful refinement of prompt engineering and the role transition of instructors into strategic mentors.

While raw final examination scores were considered, they were not utilized as a primary longitudinal metric due to substantial changes in assessment criteria and increased rigor in vocational competency standards introduced alongside the AI system. Consequently, official Student Evaluation of Teaching (SET) scores, which maintained a consistent evaluative framework, provide a more reliable longitudinal proxy for instructional effectiveness and learner satisfaction.

5.2. Student usage patterns and feature preferences

Building upon the instructional efficacy analysis presented in Section 4.2, this study further explores student usage patterns and subjective perceptions by integrating quantitative survey data (N=117) and qualitative reflections (N=35). The analysis reveals distinct behavioral preferences: in terms of pedagogical methods, 81.2% of students identified "AI Simulation Training" as their preferred approach, significantly outperforming traditional "Theoretical Lectures" (26.5%). Within the simulation environment, qualitative feedback indicates that the most utilized functions were customer conversation simulation (mentioned in approximately 80% of reflections), data analysis assistance (54%), and response generation. These preferences align closely with the anticipated scenarios identified in the structured questionnaire, such as recommending products to cross-cultural clients and managing logistics-related complaints. This congruence suggests that the AI agents successfully targeted core vocational pain points—namely high emotional load and complex data processing. Such scenarios are often difficult to replicate in conventional classrooms. Students highly valued the instantaneous nature of these interactions; for instance, one student (#35) noted that the AI's real-time analytical capabilities effectively mitigated the "brain fog" typically experienced when handling massive datasets, allowing for more focused strategic decision-making.

The efficacy of this collaborative model is fundamentally driven by the perceived realism of the simulation. Descriptive statistics show a high mean score for "Interaction Realism" at 4.03 (SD=0.67). A Pearson correlation analysis was performed to examine the relationship between perceived interaction realism and overall acceptance of the human–AI collaborative model. The analysis revealed a significant positive correlation, $r = 0.621$, $p < 0.01$ ($N = 117$). According to Cohen's (1988) conventions, this represents a moderate-to-large effect size, indicating that approximately 38.6% ($r^2 = 0.386$) of the variance in students' acceptance of the model is associated with their perceived realism of AI interactions. This finding suggests that the fidelity of the AI's role-playing is not only a critical determinant of pedagogical endorsement but also a practically significant one.

Despite these gains, students identified specific barriers that warrant attention. While overall sentiment was positive, approximately 45% of qualitative reflections described instances where AI responses felt "mechanical" or "disconnected from context," particularly in scenarios requiring deep emotional resonance. Some students (#04) critically pointed out that the AI sometimes provided excessively verbose outputs or technical glitches, such as "formatting errors and special symbols" in the exported content, which increased the burden of secondary processing. Additionally, limitations in understanding complex, multi-layered queries occasionally necessitated multiple prompt adjustments to elicit specialized depth.

6. Discussion

The findings of this study extend beyond demonstrating the effectiveness of an AI-supported instructional system. Rather, they reveal how structured human-AI collaboration reshapes learning processes in vocational training contexts. The observed improvements can be interpreted through three interconnected mechanisms: sustained experiential interaction, role-specialized cognitive scaffolding, and redistributed instructional responsibilities between human instructors and AI agents. The following discussion interprets the empirical findings from these perspectives.

6.1 Role differentiation and competency internalization

Analysis of student reflections further elucidates the psychological mechanisms behind the high pedagogical acceptance of the triadic human-AI collaborative model. Qualitative evidence indicates that students have developed a sophisticated understanding of the role differentiation between human instructors and AI agents. The AI is generally perceived as an "efficient practice partner and immediate resource repository," while the instructor is positioned as a "strategic mentor who provides value judgments and professional direction." One student feedback (#33) summarized this synergy: "AI tools offer more immediate and functional assistance for repetitive practice, while the teacher's guidance is more pertinent, providing deep insights that get straight to the point." This clear division—AI for high-frequency skill reinforcement and teachers for high-order cognitive scaffolding—validates the structural design of this study's collaborative framework.

This widely embraced model directly facilitates the internalization of professional competencies. Beyond quantitative gains, qualitative analysis reveals that AI agents empower students to transcend cognitive overloads. By offloading repetitive tasks such as initial data screening and drafting, students can reallocate their cognitive resources to higher-order strategic thinking. As noted by another student (#35), "AI customer service simulation provided solutions that allowed me to gain insights into real-world business logic, significantly enhancing my confidence in problem-solving." This sentiment reflects a fundamental shift in the learning paradigm: from traditional knowledge transmission to authentic competency development. These qualitative voices provide some clues to explain the steady increase in teaching evaluation scores observed in the quantitative data.

6.2. Implications for instructional optimization

The evolution of teaching evaluation scores—rising from 90.58 in 2022 to 93.81 in 2025—validates the current model and highlights directions for further enhancement.

1. Enhancing agent intelligence. Student feedback on "mechanical responses" and "lack of emotional resonance" (Section 4.3) indicates that future iterations must focus on emotional intelligence (EQ) and long-term dialogue memory. Expanding scenario libraries to include niche market regulations and complex crises (e.g., payment disputes or cultural taboos) can further improve instructional quality.

2. Recalibrating the instructor's role. The evaluation score increase reflects strong student endorsement of instructors as "strategic mentors." Future designs should strengthen this role by having instructors lead "feasibility debates" and "comparative case studies" after AI agents generate preliminary solutions, helping learners move from "obtaining an answer" to "internalizing business logic."

3. Evolving the assessment system. Student reflections revealed a desire for evaluations of their interaction strategies and decision-making logic, not just final output correctness. Future optimizations should use AI agent interaction data to trace student decision-making trajectories. This would enable heuristic real-time feedback when students encounter "cognitive traps," thereby supporting metacognitive development."

The transition from 'functional usability' to 'pedagogical symbiosis' identified in this study echoes the 'Human-AI Symbiosis' framework proposed by Luckin (2018), suggesting that the future of vocational training lies in an interdependent partnership where AI handles standardized cognitive loads while human instructors focus on higher-order value judgments (Luckin, 2018; Wang & Deng, 2025).

6.3 Limitations and threats to validity

Despite the encouraging empirical results, this study is subject to several limitations and potential threats to validity that should be considered when interpreting the findings. Internal validity represents a primary concern. Although the observed improvement in instructional quality coincides with the introduction of the AI Agent-based system, this study adopts a quasi-experimental design rather than a fully randomized controlled trial. A significant limitation when using longitudinal teaching evaluations is the potential for cohort effects and the "novelty effect." Different cohorts of students may possess varying baseline expectations or grading leniency when evaluating instructors. Furthermore, the initial surge in evaluation scores (from 2022 to 2023) may be partially attributed to students' excitement over interacting with novel generative AI technologies, rather than solely pedagogical improvements. Future research should employ more controlled, parallel-group experimental designs to isolate the long-term impact of the AI intervention from these confounding variables.

Construct validity is another consideration. In this study, instructional efficacy was primarily measured through student teaching evaluation scores. While these scores provide a direct indicator of student satisfaction and perceived quality, they are inherently subjective and may be influenced by the novelty effect of new technology. Although mixed-method triangulation—combining quantitative evaluations with qualitative reflections—was employed to enhance reliability, future studies could incorporate third-party professional certifications or objective behavioral analytics to further strengthen construct validity.

External validity may be constrained by the contextual specificity of the study. The system was implemented within a "Cross-Border E-Commerce Customer Management" course at a single vocational institution. While the proposed triadic framework is designed to be replicable, direct generalization to other disciplines, educational levels, or different cultural contexts should be approached with caution. Further multi-institutional studies are required to validate the broader applicability of the model across the vocational education landscape.

From a technical perspective, current AI agents still face inherent constraints. As evidenced by the empirical feedback in Section 4.3, agents based on large language models (LLMs) continue to struggle with emotional intelligence and long-term dialogue memory, leading to student reports of "mechanical" interactions. Furthermore, the "formatting errors" and "excessive verbosity" noted in qualitative reflections highlight that the system's performance remains highly dependent on the quality of prompt engineering and domain-specific knowledge coverage.

Overall, these limitations do not negate the study's findings but rather define the boundary conditions of the current system. Addressing these threats provides clear directions for future research: moving from functional usability to deeper human-machine symbiosis through multimodal integration and adaptive learning analytics.

7 Conclusions and Future Work

This study systematically addresses the practical challenges in cross-border e-commerce training by implementing a triadic "AI Agent Simulation + Teacher Collaboration" model. Empirical analysis confirms the success of this paradigm, with the average student teaching evaluation score rising from 90.58 in 2022 to 93.81 in 2025. The research validates a clear role differentiation: AI agents function as "efficient practice partners" for high-frequency skill reinforcement, while instructors serve as "strategic mentors" for value judgment and complex problem-solving.

The research demonstrates that the true value of AI in vocational education lies in fostering a more engaging, high-quality, and authentic learning environment. The consistent upward trajectory of evaluation scores—shifting from a "High Quality" baseline to a stabilized "Excellence" level (above 93.70)—serves as a compelling indicator of the success of human-AI collaborative teaching. By providing high-fidelity simulations, the system shifts the focus from passive knowledge transmission to "authentic professional readiness." This instructional optimization is further validated by the 81.2% student endorsement rate, confirming that the collaborative framework effectively meets the evolving needs of modern learners.

Looking ahead, the evolution of this model should deepen across three dimensions:

1. Technical Intelligence: Transitioning AI agents from text-based interaction to multimodal integration (speech and image) and addressing current limitations in emotional resonance to ensure more immersive training.
2. Adaptive Learning: Leveraging interaction data and learning analytics to develop personalized scaffolding that provides heuristic prompts tailored to individual student "mental blocks," thereby breaking the current evaluation plateau.
3. Eco-systemic Collaboration: Establishing a school–enterprise "Joint-Training Ecosystem" to integrate real-time industry data and niche market cases, ensuring that training scenarios evolve in tandem with the global e-commerce landscape.

In conclusion, the sustained improvement in student evaluations (from 90.58 to 93.81) marks a pivotal step toward the 'evolution and revolution' of artificial intelligence in education (Roll & Wylie, 2016). This system transforms AI into a true pedagogical partner, reshaping the digital landscape of vocational training to be more interactive and professionally relevant.

**This study was conducted in accordance with the ethical standards of Shenzhen Polytechnic University. All participants provided informed consent prior to data collection. No identifiable personal information was collected.*

Funding: This research was funded by the Shenzhen Polytechnic University 2025 Teaching Quality and Reform Project “Research on the Pathway of Integrating Generative AI into Cross-border E-commerce Customer Management Practical Training“.

References

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*, 100001.

Hwang, Y., & Lee, J. H. (2025). Exploring students' experiences and perceptions of human-AI collaboration in digital content making. *International Journal of Educational Technology in Higher Education, 22*, Article 44.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Instruction, 84*, 101733.

Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education, 4*(2), 193–212.

Lu, Y., & Tang, X. Y. (2025). Hierarchical levels and progression pathways of generative AI-empowered classroom instruction. *Research in Educational Technology, 46*(06), 75-82+106. (Original work published in Chinese)

Luckin, R. (2018). *Machine learning and human intelligence: The future of education for the 21st century*. UCL IOE Press.

Razmerita, L. (2024). Human-AI collaboration: A student-centered perspective of generative AI use in higher education. In F. Moreira (Ed.), *Proceedings of the 23rd European Conference on e-Learning (ECEL 2024)* (pp. 320–329). Academic Conferences and Publishing International.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582-599.

UNESCO. (2023). *Guidance for generative AI in education and research*. United Nations Educational, Scientific and Cultural Organization.

Wang, W., & Deng, Y. C. (2025). Symbiotic logic and practical pathways for human-machine composite educators. *Journal of Central China Normal University (Humanities and Social Sciences Edition), 64*(04), 90-99. (Original work published in Chinese)

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112.

Yi, K. Y., & Han, X. B. (2025). From blended learning to human-AI collaborative teaching: New pedagogical forms under generative AI transformation. *China Distance Education, 45*(04), 85-98. (Original work published in Chinese)

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.