Development and Empirical Study of a Multimodal Educational Agent System in Metaverse for Engineering Education Digital Transformation

Tongwei Xie
School of Business, Zhengzhou University of Technology, Zhengzhou 450044, China
*Corresponding author: 20241065@zzut.edu.cn

Abstract

Against the backdrop of digital transformation in higher education, engineering education faces critical challenges including the theory-practice divide and high barriers to technical tool adoption. This study develops a tripartite framework ("functional deconstruction-process modeling-data governance") for educational agent systems, which is built upon the Coze platform and rigorously aligned with the standards of China's Emerging Engineering Education (3E) initiative—for deployment in metaverse learning environments. By implementing multimodal interaction and adaptive decision-making mechanisms via three agent roles (teaching assistant/learning companion/management agent), we construct an immersive learning ecosystem and establish a context-aware paradigm for intelligent interactive education tailored to Chinese engineering education contexts. This research pioneers an innovative approach to cultivating engineering talent for the AI era through transformative learning methodologies.

Index Terms— Educational agent; Metaverse; Engineering education; Educational digitalization.

1 Introduction

Higher education is undergoing a pivotal phase of digital transformation, where the deep integration of artificial intelligence and metaverse technologies is fundamentally reshaping knowledge delivery paradigms and redefining the modalities of engineering education. At the national strategic level, China's Education Modernization 2035 explicitly mandates "accelerating educational reform in the information age," with joint guidelines from nine ministries emphasizing the imperative to integrate emerging technologies with pedagogical practices. Nevertheless, Chinese engineering education continues to grapple with persistent challenges, including the theory–practice dichotomy and prohibitive barriers to technical tool adoption. Specialized platforms such as MATLAB and AnyLogic—with their intricate interfaces and demanding parameter configurations—compel students to prioritize tool proficiency over domain knowledge acquisition, resulting in the paradoxical phenomenon of instrumental learning eclipsing disciplinary learning. Concurrently, prevailing online education systems—predominantly based on conventional virtual simulation platforms (e.g., standalone MATLAB simulators)—remain constrained by unidirectional interaction modes and lack persistent social interactivity, a core feature of the metaverse. This inadequacy prevents them from supporting the multifaceted experiential learning requirements of modern engineering disciplines (e.g., collaborative industrial system modeling).

While the confluence of metaverse and educational agents demonstrates transformative potential, critical research gaps persist. Current metaverse studies predominantly focus on environmental construction and immersive experiences, overlooking the mechanistic underpinnings of intelligent, personalized pedagogical agents. Internationally adopted platforms—beset by curricular and cultural mismatches with China's educational context—rely on opaque "black-box" agent models that cannot address the urgent demand for interdisciplinary talent cultivation under the Emerging Engineering Education initiative. Furthermore, existing educational agents are typically confined to single-discipline or simplified scenarios, with their capacity for dynamic coordination and conflict resolution in complex, multimodal engineering contexts (e.g., concurrent code debugging, mathematical modeling, and system simulation) remaining underexplored and unvalidated [10].

To bridge these theoretical and practical divides, this study pioneers the development of a Multimodal Educational Agent System (MEAS) within metaverse learning environments. We engineer agents with natural language interaction (supporting code-switched Chinese–English input) and adaptive guidance capabilities, and further design a dynamic coordination mechanism for multi-agent collaboration: this mechanism prioritizes task decomposition by disciplinary relevance (e.g., programming tasks assigned to teaching assistants, project scheduling to management agents) and resolves conflicts via real-time resource allocation algorithms. Additionally, we establish real-time dynamic task generation and feedback mechanisms to mitigate technical tool adoption barriers. Beyond constructing a technical framework for Multi-Agent Virtual Tutoring Systems (MAVTS), this work advances Educational Artificial Intelligence (EdAI) theory through three pivotal investigations:

1. Architecting multi-agent collaborative structures that

comprehend and adapt to complex engineering education scenarios;

2. Developing mechanisms to ensure optimality and interpretability in multi-agent task allocation and decision-making within dynamic, open-ended metaverse environments;

3. Empirically validating whether MAVTS-enhanced metaverse ecosystems significantly improve students' technical proficiency and complex problem-solving efficacy.

Through a comprehensive design–development–validation research cycle, this study delivers theoretically grounded and pragmatically actionable solutions to dismantle the learning-application disconnect in engineering education.

2 Literature Review

2.1 Current Status and Development Trends of Educational Metaverse

As a next-generation internet paradigm, the metaverse is profoundly transforming teaching and learning methodologies in education. Metaverse learning environments, characterized by their immersive, interactive, and intelligent features, offer novel pathways for educational digital transformation. Xing et al. [9] systematically delineated the application prospects of intelligent agents in the metaverse, highlighting how multimodal perception and virtual-physical fusion mechanisms extend the spatiotemporal boundaries of traditional education, particularly demonstrating pronounced advantages in engineering education domains requiring high-fidelity simulation and interactive training.

Schell et al. [6] further emphasized from a higher education transformation perspective that metaverse-enabled intelligent platforms exhibit remarkable potential in constructing authentic learning scenarios, though their pedagogical efficacy remains contingent on robust underlying intelligent technological frameworks.

Internationally, platforms such as Labster and Engage have emerged, delivering immersive experimental spaces through VR/AR technologies. However, these platforms exhibit significant misalignments with China's engineering education needs in terms of curricular systems and cultural adaptability [9].

2.2 Role Definition and Technological Evolution of Pedagogical Agents

Pedagogical agents—as core entities embodying intelligent teaching capabilities—are transitioning from conceptual frameworks to practical implementations. Wang et al. [8] synthesized in their review that pedagogical agents have evolved from early single-function proxies to comprehensive systems equipped with contextual awareness, natural interaction, and personalized decision-making capacities. Tang et al. [7] empirically validated through a 2×2 factorial experiment the interaction effects between teaching agents and affective feedback, demonstrating agents' pivotal role in enhancing user cognitive engagement [7]. Sadek et al. [5] expanded the developmental horizon of pedagogical agents along ethical dimensions through their value-sensitive design framework. Nevertheless, these studies predominantly remain confined to laboratory validations, lacking systematic evaluations in authentic educational settings.

2.3 Challenges in Integrating Metaverse and Pedagogical Agents

Despite advancements in both metaverse and pedagogical agents, their deep integration faces persistent challenges. López et al. [3] revealed through simulation studies that virtual-to-real transfer technologies have yet to resolve dynamic adaptability issues in instructional scenarios. Zhang et al. [10] identified in CAD model reconstruction research that existing agents still require improvements in decision-making precision for complex engineering tasks. Three critical gaps emerge: (1) absence of integrated architectures supporting multi-role coordination; (2) lack of solutions deeply embedded in localized teaching platforms; and (3) no universal paradigm for multimodal data processing. Pan et al. [4] substantiated through hybrid modeling that significant correlations exist between student behavioral intentions and agent functional designs, offering new insights for addressing compatibility challenges.

Building upon this research landscape, this study focuses on developing and integrating multimodal educational agents within metaverse environments, with three priority investigations: (1) architecting multi-role agent frameworks, (2) establishing rapid development pathways via the domestic Coze platform, and (3) implementing unified "functional deconstruction–process modeling–data governance" technical solutions. By systematically addressing agent adaptability, usability, and scalability in engineering education, this work aims to bridge research gaps in multifunctional integration, localized implementation, and multi-source educational data fusion—ultimately providing theoretically grounded and practically actionable paradigms for intelligent teaching systems aligned with China's higher education digital transformation.

3 Development of Educational Assistant Agents

3.1 Functional Deconstruction

This research establishes a comprehensive theoretical and technical foundation for educational agents through an integrated framework that encompasses functional deconstruction, process modeling, and data governance. As illustrated in Figure 1, the functional architecture leverages the technological ecosystem of the Coze platform to develop three distinct types of assistant agents tailored for metaverse learning environments: tutor agents, peer agents, and management agents.

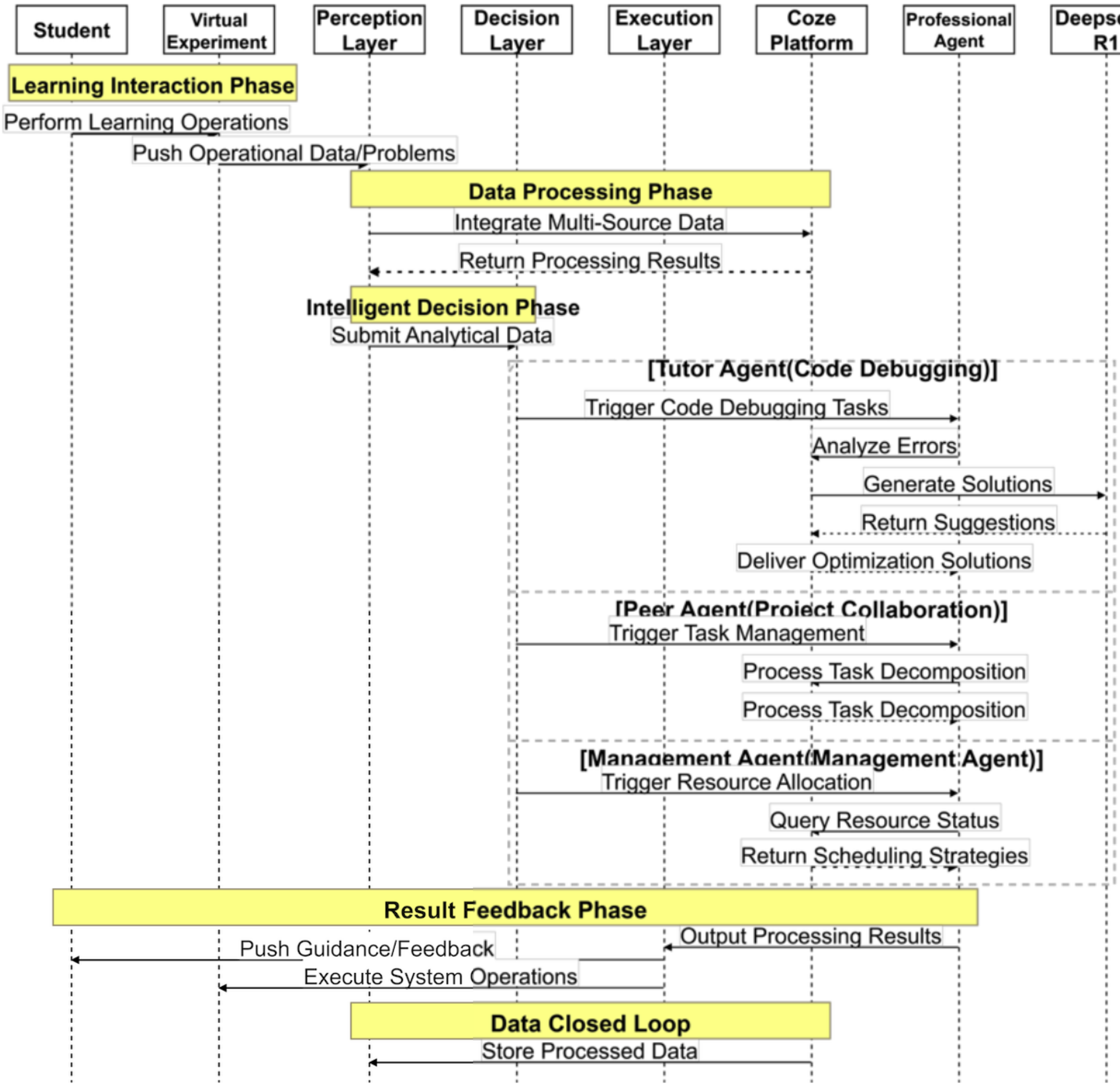

Figure 1: Functional Architecture of Educational Assistant Agents in Metaverse Learning Spaces.

The tutor agents specialize in instructional support, demonstrated in C programming courses where they identify common pointer exception errors during code submission and provide real-time debugging guidance, enabling students to efficiently modify and optimize their code. Peer agents focus on developmental guidance during project implementation, exemplified in Java Web team projects where they facilitate task decomposition, progress tracking, and conflict resolution to enhance collaborative efficiency among student groups. Management agents handle dynamic resource allocation in teaching scenarios, such as intelligently distributing computing resources for router simulators during virtual network topology experiments based on student operations, ensuring stability in multi-user concurrent learning sessions.

This precise alignment between educational objectives and agent capabilities is achieved through a dynamic role–scenario adaptation mechanism: the system first classifies learning scenarios by discipline (e.g., “Mechanical CAD 3D Modeling” vs. “Python Web Scraping”) and task type (e.g., individual debugging vs. group collaboration), then activates the corresponding agent roles (e.g., management agents for resource-intensive simulation scenarios, peer agents for collaborative projects). This mechanism ensures optimal integration between technological implementation and talent development goals.

3.2 Process Modeling

The process modeling architecture establishes a complete "perception–decision–execution" workflow chain using Coze's node-based process engine. The perception layer integrates Coze's API gateway to acquire multi-source data from virtual laboratory environments, including code compilation logs, command-line inputs, and technical queries submitted via voice or text. During voice interactions, the system utilizes Coze's built-in ASR (Automatic Speech Recognition) module to support mixed Chinese–English input. Custom plugins interface with Prometheus monitoring systems to enable comprehensive environmental data collection.

The decision layer dynamically generates learning assistance strategies by synthesizing predefined pedagogical rules with real-time interaction data. For instance, in Python web scraping courses, when over 60% of students repeatedly encounter anti-crawling mechanism errors, the agent automatically initiates targeted tutorials and case study recommendations. At the execution level, the system integrates Deepseek-R1 APIs with Coze to drive various output functions in the metaverse environment, including code demonstrations and flowchart generation, thereby completing the learning ecosystem feedback loop and ensuring immediate, context-specific instructional support.

3.3 Data Governance

The data governance framework addresses characteristic challenges of metaverse education environments, particularly data type diversity and source fragmentation, through an end-to-end management system. For structured data such as code compilation errors or graded lab reports, and unstructured data including programming design documents or group discussion recordings, the system implements schema mapping rules to establish semantic relationships. A representative application correlates compilation error types encountered in programming courses with corresponding code comments, constructing a progressive analytical framework from error classification to root-cause analysis and solution recommendations.

In data storage, we deploy a private blockchain framework (Hyperledger Fabric) to secure critical information (e.g., experimental records, skill assessment results); cryptographic hashing (SHA-256 with salting) is used for data integrity verification, and smart contracts are configured to automate access control (e.g., only instructors can modify assessment results), ensuring immutability and traceability [5]. The processing architecture employs a unified stream–batch processing approach using the Flink framework to handle dynamic data flows (e.g., real-time code submissions, student text/voice inquiries), enabling low-latency response. This system demonstrates operational efficiency in database courses, where agents can provide SQL query optimization suggestions within 900 milliseconds of detecting student difficulties, significantly accelerating technical skill acquisition.

4 Application of Educational Assistant Agents in Metaverse Learning Spaces

The effective deployment of intelligent agents within metaverse learning environments fundamentally relies on multi-platform integration and ecosystem construction. As the pilot institution for this study, Zhengzhou University of Technology has established a Unity3D-based metaverse learning infrastructure that serves as the foundation for our implementation. Building upon open interfaces, we have developed a lightweight SDK and API gateway that encapsulate core functionalities including knowledge retrieval and interactive response mechanisms, thereby achieving seamless interoperability between our agent system and virtual simulation platforms [1].

As demonstrated in Figure 2, the implemented system enables agents to access and utilize existing experimental resources within the metaverse learning environment through several key integrations: bidirectional data exchange with the Moodle learning management system facilitates automated assessment of student programming exercises; direct interfacing with MATLAB's online environment supports mathematical modeling activities; and comprehensive guidance mechanisms assist students in completing algorithm verification tasks within virtual scenarios.

To ensure continuous adaptation to evolving metaverse learning spaces, we have designed configurable functional modules tailored to specific engineering disciplines. For instance, in the Smart Manufacturing program, which emphasizes hands-on operational skills, we have implemented enhanced weighting for 3D modeling guidance within Mechanical CAD courses. Conversely, for Computer Science disciplines that prioritize programming application competencies, additional code debugging assistance modules have been incorporated into Python courses. This discipline-specific customization achieves precise alignment between agent capabilities and professional training objectives.

In summary, through multi-platform integration, ecosystem development, and continuous adaptation, our educational assistant agent system addresses critical pedagogical needs in metaverse learning environments, including knowledge guidance and natural interaction support. This implementation establishes an operational paradigm for the digital transformation of engineering education that effectively bridges technological innovation with practical teaching requirements.

5 Methodology

5.1 Research Design and Experimental Protocol

This study employs a quasi-experimental research design with control and experimental groups to systematically evaluate the impact of educational agent systems on engineering students'

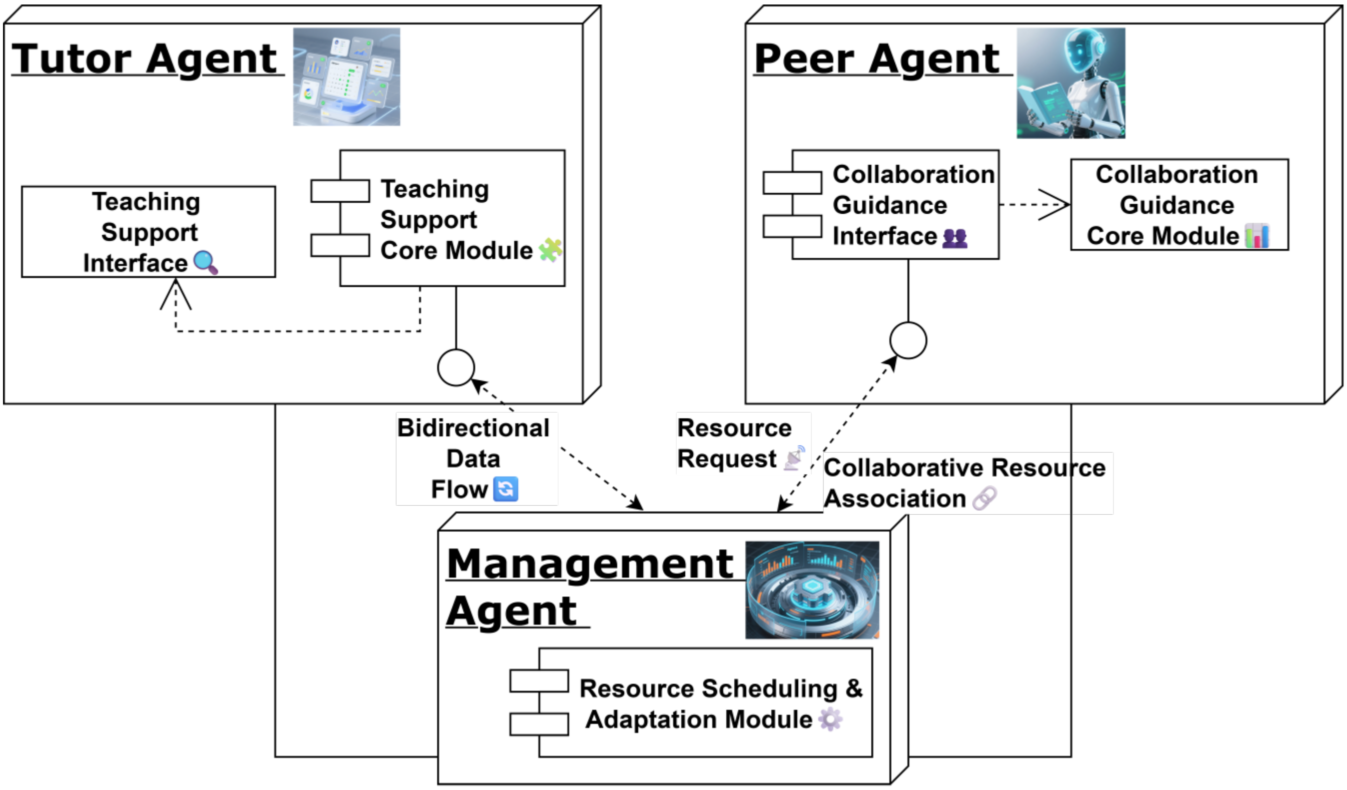

Figure 2: Integration and Application Scenarios of Tutor–Peer–Management Agents.

learning outcomes in metaverse environments. Conducted at Zhengzhou University of Technology, the experiment recruited 60 engineering majors from the 2024 cohort as participants. A simple random sampling method based on odd–even student ID stratification was implemented, and random assignment was used to form two groups: the experimental group (30 students) utilized a metaverse learning system integrated with multimodal educational agents, while the control group (30 students) employed conventional university-level virtual simulation platforms. Both groups followed an identical Systems Engineering curriculum.

To ensure internal validity, pre-intervention assessments using the ACM–ICPC elementary question bank confirmed no significant baseline difference between groups (t = 0.275, p = 0.784), satisfying the homogeneity assumption for group comparisons [4]. The 8-week intervention maintained rigorous control over instructional pacing, content delivery, and instructor variables to minimize confounding effects.

5.2 Measurement Instruments and Data Collection Framework

The study implemented standardized measurement tools with established reliability and validity, comprising two core instruments: the Technical Tool Proficiency Assessment Scale and the Programming Task Efficiency Evaluation Protocol. The 21-item proficiency scale measures three dimensions: engineering software operation, parameter configuration/debugging, and comprehensive application. The efficiency protocol quantifies task completion accuracy, time expenditure, and code quality for specified programming assignments under timed conditions.

For the experimental group, embedded data acquisition modules within the metaverse platform captured real-time multimodal interaction logs, including voice-based Q&A transcripts, code submission sequences, and task response pathways. All data collection followed double-blind procedures, with research assistants uninvolved in the experiment performing data cleansing and anonymization to ensure objectivity and regulatory compliance.

6 Experimental Validation

6.1 Psychometric Evaluation of Measurement Instruments

A comprehensive validation of research instruments was conducted to ensure reliability. For construct validity, the Kaiser–Meyer–Olkin (KMO) measure yielded 0.872, and Bartlett's test of sphericity produced $\chi^2 = 326.417$ (df=21, p<0.001), confirming the data's suitability for factor analysis. Principal component analysis extracted three factors with eigenvalues exceeding 1.0, collectively explaining 82.3% of cumulative variance. All items demonstrated factor loadings ranging from 0.612 to 0.832 without cross-loading, indicating excellent discriminant validity. Reliability analysis revealed strong internal consistency, with an overall Cronbach's α coefficient of 0.891

and subscale coefficients between 0.802 and 0.856—all surpassing the 0.7 acceptability threshold. Complete metrics are presented in Table 1.

6.2 Learning Outcome Comparative Analysis

Independent samples t-tests were employed to assess the educational agents' impact on learning efficacy. Results (Table 1) demonstrated the experimental group's significant superiority over controls in both technical tool proficiency (84.6±5.7 vs. 73.2±8.4, t=5.892, p=0.003) and programming task efficiency (86.3±4.9 vs. 70.8±9.1, t=7.341, p=0.001). To account for pre-test variations, ANCOVA with pre-test scores as covariates maintained statistical significance ($F(157) = 32.17$, p<0.001), confirming the robust positive effect of agent intervention on learning outcomes.

6.3 Interaction Behavior and Mechanistic Analysis

Behavioral sequence analysis of experimental group interaction logs revealed a 42% reduction in average error resolution time and 35% improvement in code iteration efficiency among agent-assisted students. Semantic mining identified two key agent functionalities: real-time error diagnosis (e.g., pointer exception debugging, SQL query optimization) and personalized strategy delivery (e.g., targeted anti-crawling mechanism tutorials). These findings demonstrate that educational agents serve as precision "cognitive scaffolding" through multimodal interaction and adaptive decision-making mechanisms, effectively optimizing the learning process.

7 Discussion

7.1 Interpretation of Results and Mechanistic Insights

The experimental results demonstrate that the metaverse learning environment integrated with educational agents significantly enhances engineering students' technical tool proficiency and programming task efficiency. This outcome can be explained through two key mechanisms.

First, the multimodal interaction framework of educational agents—encompassing natural language queries, real-time feedback, and dynamic task generation—provides highly contextualized and personalized learning guidance. Prior studies have shown that such interaction patterns effectively reduce the operational complexity of engineering tools, mitigating the "tool learning overshadowing disciplinary learning" phenomenon, as discussed by Wang et al. [8]. Tang et al. [7] further corroborate these efficiency gains through empirical evidence demonstrating the synergistic effects of pedagogical agents and affective feedback.

Second, the integration of immersive simulation scenarios with agents' adaptive decision-making capabilities creates an integrated "learn–practice–evaluate" experience. López et al. [3] highlight how virtual–physical hybrid environments markedly enhance knowledge transformation, a finding that aligns with the behavioral performance improvements observed in our study. Behavioral data analysis shows that experimental group students exhibited 42% shorter error resolution times and 35% higher code iteration efficiency—results consistent with Hu et al. [2], whose work demonstrates the effectiveness of conversational agents in supporting collaborative learning.

7.2 Comparative Analysis with Existing Research

Compared to international studies, our educational agent system demonstrates distinct advantages in functional integration and scenario adaptability. Schell et al. [6] identify limitations in mainstream platforms' intelligent interaction capabilities, whereas our "tutor–peer–management" multi-agent architecture provides comprehensive instructional support. This design aligns with the integrated metaverse agent system vision proposed by Xing et al. [9].

Relative to domestic research, Zhang et al. [10] focus primarily on domain-specific CAD reconstruction, while our process modeling and data governance approaches enable cross-scenario agent coordination. In addition, Li et al. [2] report varied cognitive load impacts of pedagogical agents; our mixed-methods evaluation (quantitative analysis combined with behavioral sequence analysis) provides more robust evidence supporting these findings.

From a technological standpoint, Hu et al. [1] propose an edge–cloud hybrid architecture for AI agents, which informed part of our system implementation. Pan et al. [4] demonstrate the value of SEM–ANN methodology for educational agent evaluation, and these insights guided our experimental design. Furthermore, Sadek et al. [5] critique single-solution paradigms in agent design; our integrative multi-agent architecture addresses these limitations and establishes new benchmarks for functional completeness and contextual adaptability.

8 Conclusion

This study has conducted practical research on the development, application, and integration of multimodal educational agent systems within metaverse learning environments, specifically tailored to the pedagogical characteristics of engineering and software-related courses in higher education. By establishing an intelligent interactive education paradigm and immersive learning ecosystem, we have addressed two fundamental challenges in traditional engineering education: the prohibitive barriers to technical tool adoption and the persistent disconnect between theory and practice. Our system enhances the application of theoretical knowledge in practical tasks, providing students with more actionable learning experiences.

However, several limitations must be acknowledged. The current implementation scope remains confined to selected

Table 1: Comparative Analysis of Test Data Between Experimental and Control Groups

Test Category	Metric	Experimental Group	Control Group	Statistical Value	p-value	Threshold	Result
Pre-test Balance	Pre-test Score	72.4±6.8	71.9±7.2	t=0.275	0.784	>0.05	No Significant Difference
KMO Test	Sampling Adequacy	–	–	0.872	–	0.872	Excellent Fit
Bartlett's Test	Sphericity Test	–	–	$\chi^2 = 326.417$, df=21	<0.001	<0.05	Significant
Factor Analysis	Cumulative Variance Explained	–	–	82.3%	–	>60%	Outstanding
	Item Communality Range	–	–	0.612–0.832	–	>0.4	Ideal
Reliability Test	Cronbach's α	–	–	0.891	–	>0.7	Excellent
	Subscale α	–	–	0.802–0.856	–	>0.7	Good
Post-test Performance	Technical Tool Proficiency	84.6±5.7	73.2±8.4	t=5.892	0.003	<0.05	Significant Difference
	Programming Task Efficiency	86.3±4.9	70.8±9.1	t=7.341	0.001	<0.01	Highly Significant Difference

engineering disciplines at pilot institutions, necessitating future expansion to diverse academic programs across national higher education systems to validate the generalizability of our findings. Furthermore, the long-term effects of agent-assisted learning—particularly the underlying mechanisms through which these agents influence the development of engineering thinking—require sustained longitudinal investigation and deeper exploration.

Looking ahead, as metaverse technologies continue to mature, more immersive agent-assisted learning modalities are poised to become a pivotal direction for the digital transformation of higher education. From a technological evolution perspective, breakthroughs in multimodal interaction could enable agents with enhanced affective computing capabilities, such as precisely capturing students' cognitive states during virtual engineering experiments via facial expression analysis and voice pattern recognition. These capabilities would allow for dynamic instructional adaptation, delivering optimally tailored guidance and motivational support.

In terms of application expansion, future advancements in agent collaboration may achieve qualitative leaps—from supporting single-scenario interactions to enabling multi-scenario coordination and facilitating synchronous multi-learner collaboration within shared virtual spaces—ultimately realizing seamless knowledge interconnection across metaverse environments.

As a critical nexus between virtual and physical learning spaces, and between instructors and learners, educational agents will catalyze the transformation of higher education from standardized training to precision education. This paradigm shift will forge robust couplings among educational ecosystems, talent pipelines, and industrial needs, providing a solid human capital foundation for building an innovation-driven nation.

Funding

Supported by the Construct Program of the Key Discipline in Zhengzhou University of Technology. Key Scientific Research Project of Henan Provincial Institutions of Higher Education in 2026: "Research on Collaborative Innovation Mechanism and High-Quality Development Path of University–Enterprise R&D Centers in Henan Province Driven by Dual Wheels" (Project No.: 26B630028).

References

[1] Liao Hu, Qiteng Wu, and Ruoyu Qi. Empowering smart app development with solidgpt: an edge–cloud hybrid ai agent framework. *Advances in Engineering Innovation*, 16(7):119–125, 2025.

[2] Wanqing Hu, Rushi Gong, Sisi Wu, and Yanyan Li. A conversational agent based on contingent teaching model to support collaborative learning activities: impacts on students' learning performance, self-efficacy and perceptions. *Educational Technology Research and Development*, pages 1–32, 2025. Prepublish.

[3] Lucía Güitta López, Lionel Güitta López, Jaime Boal, and Álvaro J. López López. Sim-to-real transfer via a style-identified cycle consistent generative adversarial network: Zero-shot deployment on robotic manipulators through visual domain adaptation. *Engineering Applications of Artificial Intelligence*, 159:111510–111510, 2025.

[4] Yingying Pan, Jingyang Zhang, Jiahui Yu, and Zhehao Pan. Modeling primary school students' behavioral intention to use genai-augmented pedagogical agents: a hybrid sem-ann method. *Education and Information Technologies*, pages 1–20, 2025. Prepublish.

[5] Malak Sadek, Rafael A. Calvo, and Céline Mougenot. The value-sensitive conversational agent co-design framework. *International Journal of Human–Computer Interaction*, 41(15):9533–9564, 2025.

[6] Julie Schell, Kasey Ford, and Arthur B. Markman. Building responsible ai chatbot platforms in higher education: an evidence-based framework from design to implementation. *Frontiers in Education*, 10:1604934–1604934, 2025.

[7] Xiaoqin Tang, Li Jiang, Guoli Liu, and Hongxia Li. The interaction effect of pedagogical agent and emotional feedback on effective learning: a 2 × 2 factorial experiment in online formative assessment. *Frontiers in Psychology*, 16:1610550–1610550, 2025.

[8] Yanlin Wang, Wanjun Zhong, Yanxian Huang, Ensheng Shi, Min Yang, Jiachi Chen, and Zibin Zheng. Agents in software engineering: survey, landscape, and vision. *Automated Software Engineering*, 32(2):70–70, 2025.

[9] Luran Xing, Zhengping Li, and Lijun Wang. Applications and future trends of agents in the metaverse. *SID Symposium Digest of Technical Papers*, 56(S1):951–954, 2025.

[10] Chao Zhang, Arnaud Polette, Romain Pinquié, Mirai Iida, Henri De Charnace, and Jean Philippe Pernot. Reinforcement learning-based parametric cad models reconstruction from 2d orthographic drawings. *Computer-Aided Design*, 188:103925–103925, 2025.