

Research on the Impact of Virtual Digital Human Teachers on the Interactivity of Aesthetic Education Teaching

Yawei Zhang*

Zhejiang Yuexiu University of Foreign Languages, Shaoxing, 312000, China

*Corresponding author: 20152022@zyufl.edu.cn

Abstract: With the deep integration of artificial intelligence, virtual reality, and multimodal interaction technologies, the application of virtual digital human teachers in the field of education has shown a trend of evolution from auxiliary tools to intelligent interactive agents, particularly in aesthetic education, which emphasizes aesthetic experience and emotional resonance, where their unique value has become increasingly prominent. This paper conducts a systematic study on the impact of virtual digital human teachers on the interactivity of aesthetic education teaching, analyzing their technological connotations, interaction characteristics, and cognitive mechanisms, and clarifying the generative logic and evolutionary path of virtual teaching agents. On this basis, it explores their aesthetic expression, human-computer emotional regulation, and “human-likeness” acceptance mechanism within teaching interaction systems, and further proposes teaching structure optimization strategies based on aesthetic configuration and knowledge reorganization. By constructing multidimensional interaction models and process-oriented evaluation systems, this paper attempts to outline the overall framework and future development of the involvement of virtual digital humans in aesthetic education teaching, aiming to provide theoretical support and methodological references for enhancing art teaching interaction in digital education environments.

Keywords: virtual digital human teachers; aesthetic education teaching; human-computer interaction; aesthetic experience; interaction structure

Introduction

Against the backdrop of the continuous advancement of educational intelligence, aesthetic education teaching, as a complex teaching domain integrating perceptual experience, emotional expression, and artistic understanding, places higher demands on the interactivity and contextuality of teaching media. Traditional teaching models face multiple limitations in content delivery and emotional co-construction, whereas virtual digital human teachers, with their humanized expression, multimodal perception, and semantic intelligence capabilities, are gradually becoming an important technological force in reshaping the logic of aesthetic education interaction. Through multidimensional perception systems and aesthetic configuration capabilities, they can enhance aesthetic resonance and expressive synergy between teachers and students, thereby promoting the transformation of teaching relationships toward emotional connection and co-creation of meaning. This paper focuses on the in-depth participation mechanisms of virtual digital humans in aesthetic education teaching, aiming to conduct a systematic discussion from the perspectives of interaction technology, human-computer relationships, and teaching structures, to analyze their impact mechanisms on aesthetic interaction, emotional regulation, and cognitive acceptance, and to propose teaching structure optimization paths and interaction quality evaluation models, in order to provide theoretical support and technical insights for the transformation of aesthetic education teaching in the era of digital intelligence.

1. Definition and Development Logic of Virtual Digital Human Teachers

1.1 Interpretation of the Connotation and Technological Support of Virtual Digital Human Teachers

A virtual digital human teacher is a digital educational agent that integrates artificial intelligence, semantic modeling, human-computer interaction, multimodal perception, and dynamic simulation, and possesses autonomous cognition, semantic understanding, and multisensory expression capabilities. Its

core feature lies in simulating, through algorithmic models, the linguistic logic, interaction strategies, and emotional expressions of human teachers in educational settings, thereby constructing a digital identity endowed with both teaching functions and emotional attributes. In terms of generative mechanisms, a virtual digital human relies on natural language processing technologies within deep learning frameworks to achieve semantic generation, while combining three-dimensional modeling and speech synthesis systems to produce dynamic facial expressions and intonation adjustments, ensuring the fluency and naturalness of human–computer interaction. Furthermore, through learner data analysis and adaptive learning models, a virtual teacher can continuously optimize its output content and interaction pace based on individual feedback, thereby achieving the dynamic generation of personalized teaching strategies ^[1].

From the perspective of technological evolution, the development of virtual digital human teachers is not a simple accumulation of technologies but the result of collaborative advancement across multiple frontier fields. Computer graphics provides visual support for image construction, while neural networks and multimodal fusion models endow them with “human-like” cognitive and contextual adaptation capabilities. Meanwhile, data annotation and behavior modeling based on cognitive science and educational psychology transform the teaching behaviors of virtual digital human teachers from passive responses into dynamic processes that are predictable, controllable, and capable of evolution. Compared with traditional digital educational tools, their greatest breakthrough lies in breaking the static boundary of information transmission by introducing a teaching mechanism of “intelligent interaction + emotional adaptation,” thus demonstrating significant advantages in constructing a sense of realism, interactivity, and immersion in teaching.

1.2 Dimensions of Interactivity and Aesthetic Characteristics in Aesthetic Education Teaching

Aesthetic education teaching is not merely a process of knowledge transmission but also a multidimensional construction system for generating aesthetic experience, facilitating emotional exchange, and achieving value resonance. Its interactivity is reflected in the symbolic exchange and emotional connection between teachers and students on the basis of aesthetic experience, resulting from the co-construction of cognition, perception, and emotion. In the teaching context, interactivity is not only manifested in verbal responses and feedback but also involves the reception and reproduction of nonverbal symbols—such as visual images, music, and movement—forming a complex multichannel communication system. The unique artistic context of aesthetic education makes “interaction” more than an informational dialogue; it becomes an empathetic connection centered on aesthetic perception, requiring teaching agents to possess the ability to stimulate aesthetic sensitivity, guide artistic interpretation, and organize aesthetic expression, thereby enabling students to construct meaning through immersion and expand understanding through dialogue.

The characteristics of interactivity in aesthetic education lie in its high openness and contextual dependence, differing from the interaction patterns in science education, which emphasize logical rigor and standard answers. Its interaction mechanisms are often embedded within visual and emotional cognition, and teaching effectiveness depends on individuals’ thresholds of aesthetic perception and experiential background. In this process, artworks are no longer unidirectional carriers of cultural information but serve as media for the flow of emotions and the construction of meaning among multiple agents. When a virtual digital human acts as a teaching agent in this context, its mode of interaction must not only possess technical responsiveness and semantic comprehension but also reflect depth in artistic perception and flexibility in aesthetic expression. Therefore, the interactivity of aesthetic education teaching is not merely a technical issue but also a manifestation of the complex relationship between aesthetic experience and cultural understanding ^[2].

1.3 Evolutionary Path of Virtual Digital Humans and Teaching Interaction Modes

With the continuous integration of artificial intelligence and virtual reality technologies, the role of virtual digital humans in teaching contexts has undergone a structural leap from “tool-based presentation” to “participatory subject.” In the early stages, virtual images were mainly used for visualizing teaching resources, with their interactive dimension limited to playing preset content and providing command-based feedback, lacking dynamic generation and emotional comprehension capabilities. In the intelligent stage, with the improvement of deep learning and multimodal perception capabilities, virtual digital human teachers have gradually acquired complex abilities such as contextual perception, intent recognition, and emotional interaction, enabling them to make real-time adjustments based on learners’ language input, facial expressions, and behavioral responses, and to generate

teaching feedback aligned with the given context. Particularly in the field of aesthetic education, this evolved interaction model breaks through the linear structure of traditional classrooms, infusing the teaching process with openness and creativity, and facilitating the stimulation of learners' active participation and capacity for artistic re-creation.

At present, the interaction path of virtual digital humans in teaching is moving toward "collaborative intelligent interaction," characterized by a tripartite co-construction of the teaching ecosystem by teachers, students, and virtual intelligent agents. In this structure, virtual digital humans not only imitate human teaching behaviors but also assume multiple roles as "co-creators of teaching" and "artistic facilitators." Their interaction design must balance teaching objectives, aesthetic contexts, and individual differences, promoting the evolution of interaction behaviors from "command-response" to "perception-co-construction." This shift not only enhances the layering and flexibility of teaching interaction but also reconstructs the teacher-student relationship and knowledge transmission logic in aesthetic education. In the future, with the further development of semantic computing, emotional modeling, and cultural embedding mechanisms, virtual digital humans are expected to participate at a higher level in aesthetic judgment, creative stimulation, and artistic reflection, thereby continuously innovating the deep interactive structures of aesthetic education teaching [3].

2. Construction Logic of the Interaction Mechanism of Virtual Digital Humans in Aesthetic Education Teaching

2.1 Aesthetic Expression Modes in the Human-Computer Interaction Domain

In aesthetic education teaching involving virtual digital humans, human-computer interaction is not only a process of information transmission but also a complex construction of aesthetic behavior. Aesthetic expression, as the core dimension of interaction, manifests as the multidimensional coordination of visual perception, language style, emotional regulation, and cultural metaphor. Virtual digital humans simulate and extend the expressive capabilities of human teachers in art teaching through artistic stylization in character design, rhythmic aesthetics in voice tone and timbre, and synchronized body language. Their interactive contexts create an aesthetically rich and immersive learning environment via dynamic response mechanisms and scene construction algorithms, enabling students to generate aesthetic judgments, experience artistic influence, and activate emotional resonance through multisensory reception.

Multimodal interaction constitutes the key path for virtual digital humans to realize aesthetic expression, centering on the unified integration of linguistic, visual, motor, and auditory modal information into semantically coherent aesthetic outputs. In aesthetic education classrooms, virtual teachers can incorporate elements of visual art, performing arts, and language arts by mobilizing facial expressions, gestures, and accompanying music, thereby achieving rich aesthetic semantic expression. Compared with static teaching resources, such integrated expression offers greater contextual rendering and emotional guidance effectiveness, effectively stimulating students' perceptual depth and associative breadth regarding artistic information. Through algorithm-driven dynamic composition and contextual interpretation, virtual digital humans do not merely transmit aesthetic knowledge but actively participate in the generation and negotiation of aesthetic experience [4].

2.2 Technological Intervention of Virtual Digital Humans in Learning Motivation and Emotional Regulation

In aesthetic education teaching, students' learning motivation and emotional states significantly affect teaching outcomes. Virtual digital humans utilize contextual perception technology and behavioral data capture capabilities to dynamically identify students' attention levels, engagement enthusiasm, and emotional fluctuations during teaching, enabling precise pedagogical intervention and guidance. By integrating emotion recognition algorithms with learning state analysis models, the system can fuse multidimensional signals such as learners' behavioral performance, vocal tone, and eye movement trajectories, and adjust teaching tone, feedback methods, and presentation content in real time to enhance the emotional adaptability of teaching interaction. This technological intervention mechanism not only helps maintain students' focus and learning investment but also builds a more empathetic teaching experience in the highly emotional domain of aesthetic education.

Beyond perception and adaptation functions, virtual digital humans can guide and motivate students' learning motivation through behavior strategy modeling. For example, in art creation tasks, the system

can generate customized positive feedback based on student performance to enhance their self-efficacy and aesthetic confidence. Meanwhile, through humanized emotional resonance and positive verbal encouragement mechanisms, virtual teachers can construct supportive teaching relationships within a psychologically low-pressure environment, thereby promoting students' exploration of artistic expression and emotional projection possibilities in a safe atmosphere. This process not only represents technological intervention in emotional regulation but also signifies the deep value of human-computer interaction in stimulating educational motivation and regulating the learning process.

2.3 Cognitive Acceptance Mechanism of the “Human-likeness” Feature in Aesthetic Interaction

“Human-likeness” constitutes the fundamental prerequisite for virtual digital humans to establish effective teaching interactions, especially in aesthetic education, where the affinity and trust it generates significantly influence students' emotional investment and cognitive acceptance. Human-likeness is not limited to physical resemblance but also involves the naturalness of verbal logic, authenticity of emotional expression, and contextual relevance of behavioral responses. Learners, during interaction with virtual teachers, construct a cognitive representation of the teacher's “social identity” based on language style, facial expressions, and feedback mechanisms, which determines whether the virtual teacher is integrated into the category of real teaching agents. This process involves social attribution mechanisms in cognitive psychology, where individuals assign social roles and interaction intentions to virtual entities based on symbolic and behavioral cues during interaction [5].

If virtual teachers in aesthetic education effectively activate the “human-likeness cognitive pathway,” they can establish deep interactive relationships without physical contact. In teaching interaction, human-likeness helps reduce students' cognitive distance and increase their acceptance of digital teachers' instructions and emotional outputs. When the system exhibits human-like microexpressions, semantic logic, and appropriate emotional responses, students are more likely to perceive it as a trustworthy teaching partner rather than a cold technical tool. This acceptance mechanism based on human-likeness cognition forms the foundational framework for emotional co-construction and aesthetic value collaboration in aesthetic interaction, and provides cognitive and psychological legitimacy for the deep participation of virtual digital human teachers.

3. Optimization Paths for the Interaction Structure of Aesthetic Education Teaching Involving Virtual Digital Humans

3.1 Aesthetic Configuration of Interaction Content and Knowledge Reorganization

In aesthetic education teaching involving virtual digital humans, the design of interaction content should not be limited to the linear presentation of traditional knowledge units but should center on aesthetic experience, reconstructing the structural logic and forms of expression of teaching content. Based on the openness characteristic of the art ontology, knowledge in the context of aesthetic education needs to be encoded in ways that are contextualized, symbolic, and polysemous, thereby guiding learners to generate personalized aesthetic associations and cultural interpretations through interaction. Virtual digital humans can aesthetically reconstruct original teaching content through semantic enhancement, multimodal interpretation, and image-language collaboration techniques, transforming knowledge into interactive objects endowed with emotional appeal and artistic tension. This content construction path emphasizes generativity of meaning and multidimensional cognition, providing fundamental support for achieving deep teaching interaction [6].

The knowledge reorganization process involves the visual semantic restructuring and cognitive rhythm regulation of aesthetic education content. Virtual digital human teachers can dynamically adjust the density of content presentation, aesthetic styles, and symbol combinations based on learner feedback and changes in cognitive load, thus realizing personalized and progressive teaching advancement. Meanwhile, their algorithm-driven content generation capability enables the embedding of artistic knowledge into everyday life contexts and cross-media expressions, constructing multilayered and multipath knowledge network structures, thereby breaking the linear boundaries of knowledge transmission in traditional classrooms. Through this reorganization strategy centered on aesthetic configuration, teaching content achieves higher levels of contextual expressiveness and aesthetic appeal while meeting disciplinary norms, reinforcing the generativity and fluidity of knowledge in virtual interactions.

3.2 Generation Mechanism of Interaction Strategies in the Teaching Context

The interaction strategies employed by virtual digital humans in aesthetic education teaching require establishing a dynamic coordination mechanism among teaching objectives, student characteristics, and artistic contexts. The generation mechanism of interaction strategies depends on the virtual teacher's real-time recognition and semantic reasoning abilities regarding the teaching context, including critical processes such as emotional state analysis, contextual adaptation, and interaction intention prediction. Within the emotionally oriented and flexibly expressive discipline of art education, virtual digital humans need to construct appropriate dialogue models and feedback patterns based on the teaching situation, ensuring that the interaction process possesses the expressive power of artistic language and the depth of emotional engagement. The system forms an interaction cycle of guidance—response—re-guidance by imitating human teachers' guiding language, symbolic cues, and evaluative responses, continuously stimulating learners' emotional resonance and expressive desire.

The core of virtual interaction strategies lies not in the efficiency of information transmission but in the construction and maintenance of an aesthetic co-creation relationship. By introducing context-aware strategy optimization models, virtual teachers can generate diversified teaching response paths based on students' interaction behavior characteristics—such as pause frequency, emotional fluctuations, and language style—including modes of encouragement, guidance, questioning, and extension, thereby improving the plasticity of interaction and adaptability to the teaching context. Supported by interaction data, this mechanism achieves dynamic iteration of strategies through behavioral reasoning and intention recognition algorithms, rendering the teaching interaction process nonlinear, decentralized, and co-constructed, fully embodying the requirements of openness, collaboration, and creativity in aesthetic education teaching.

3.3 Evaluation Dimensions of Teaching Interaction Systems Involving Virtual Digital Humans

The deep involvement of virtual digital humans in the interaction structure of aesthetic education teaching prompts a transition of teaching interaction systems from static evaluation to dynamic, multidimensional, and process-oriented evaluation models. Traditional teaching effectiveness assessments often focus on explicit indicators such as knowledge mastery and behavioral performance, whereas in virtual interaction contexts, greater attention is needed on implicit dimensions including learners' emotional investment, symbolic interpretation ability, and artistic expressive tension during aesthetic interaction. Evaluation of interaction systems should establish a comprehensive indicator system encompassing emotional responsiveness, aesthetic participation, expressive generativity, and interaction adaptability, extracting key characteristic variables from dynamic processes and multimodal feedback to enable in-depth analysis of virtual interaction quality.

Based on algorithm-driven process analysis mechanisms, the system can collect and analyze interaction data in real time during teaching, including linguistic content, facial expressions, speech rate changes, and bodily movements, thus forming a multidimensional and multigranular interaction evaluation model. This evaluation mechanism not only provides precise evidence for optimizing teaching strategies but also supplies feedback for the evolution of virtual digital humans' capabilities. In the field of aesthetic education, this mechanism helps identify learners' developmental trajectories in artistic expression, thereby constructing an aesthetic development assessment framework centered on experience and guided by feedback. Through the combination of refined interaction quality evaluation and generative teaching feedback, the teaching system achieves a leap from “teaching—learning” to “co-creation—collaboration,” promoting the sustained evolution of the virtual aesthetic education teaching ecosystem.

Conclusion

The introduction of virtual digital human teachers provides a new opportunity for the reconstruction of the interaction structure in aesthetic education teaching. Their expressive adaptability, emotional resonance, and strategy generation capabilities demonstrated within the aesthetic context significantly expand the boundaries of traditional teaching interaction. Research shows that virtual teachers possess a high potential for synergy in multimodal interaction, emotional regulation, and aesthetic knowledge configuration, effectively enhancing students' engagement and artistic expressive abilities, thereby constructing a more immersive and humanized aesthetic learning experience. In the future, the deep integration of virtual digital humans in teaching scenarios still requires breakthroughs in overcoming

the contextual limitations of semantic understanding, precision bottlenecks in emotional modeling, and personalization issues in cognitive adaptation. At the level of teaching system design, it is urgent to build interactive architectures with more dynamic feedback mechanisms and generative support to realize a new ecosystem of co-creative art education between humans and machines. Based on this, the virtual digital human teaching interaction paths and evaluation models proposed in this paper not only provide operational guidance for the current stage of digital aesthetic education teaching but also offer an expansion space for future research on ethical design, cultural adaptation, and aesthetic evaluation of intelligent educational agents.

References

- [1] Chen Shi. "Analysis of the Current Development and Risk Countermeasures of Virtual Digital Humans." *Broadcasting and Television Network* 32.06 (2025): 34–38.
- [2] Tang Qing. "Teaching Practice Research on the Integration of Virtual Digital Human Technology and Art from the Perspective of the Metaverse." *Toy World* 05 (2025): 215–217.
- [3] Lu Wenzhong. "Innovative Exploration of Virtual Digital Human Integration into Higher Education Teaching." *Journal of Multimedia and Network Teaching in China (Early Edition)* 03 (2025): 25–28.
- [4] Chen Hongcai. "Application of Virtual Digital Human Technology in Information Interaction on Online Education Platforms." *Wireless Internet Technology* 21.24 (2024): 68–71.
- [5] Yang Heming. "Design and Application of Virtual Digital Humans Based on Deep Learning." *Electronics Technology* 53.11 (2024): 60–61.
- [6] Jiang Ting. "Innovative Applications and Strategic Research of Virtual Digital Human Technology in Digital Education." *Science and Technology Information* 22.14 (2024): 211–214+219.