

# Advancements in Digital Humans for Recorded Courses: Enhancing Learning Experiences via Personalized Interaction

Qi Liu<sup>a</sup>, Yunhao Sha<sup>a</sup>, Kai Zhang<sup>b</sup>, Zhenya Huang<sup>b</sup>, Linbo Zhu<sup>b,c</sup>, Junyu Lu<sup>a,c</sup>, Yu Su<sup>c,d</sup>

<sup>a</sup> School of Artificial Intelligence and Data Science, University of Science and Technology of China, Hefei 230026, China

<sup>b</sup> School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China

<sup>c</sup> Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

<sup>d</sup> School of Computer Science and Artificial Intelligence, Hefei Normal University, Hefei 230061, China

© Higher Education Press 2025

**Abstract** In recent years, significant advancements in educational technology, particularly the rise of online learning platforms, have transformed the way students engage with educational content. Most students currently access learning materials through live streaming or pre-recorded courses, but live sessions are often confronted with scheduling conflicts between instructors and students, thereby limiting their accessibility. This compels numerous students to opt for pre-recorded lessons, which, despite offering flexibility, lack real-time interaction and direct guidance. The unavailability of professional educators creates a challenge for parents, who are frequently unable to assist their children effectively with academic difficulties. Consequently, when students encounter obstacles during self-paced learning involving pre-recorded materials, their motivation to continue diminishes considerably. To address these challenges, this study developed an intelligent digital human-based teaching system that leverages the capabilities of large language models to provide personalized, interactive, and real-time support. It is designed to simulate the presence of a human instructor, enhancing the learning experience by offering tailored guidance, answering queries, and adapting to individual progress, features that ultimately foster a more engaging and efficient educational process. The proposed solution seeks to bridge the gap between the flexibility of pre-recorded lessons and the interactive, instantaneous engagement typically associated with live courses.

**Keywords** digital human, pre-recorded courses, large language models (LLMs)

## 1 Introduction

The past few years have seen the demand for high-quality, scalable, and personalized education drive the rapid development of digital learning systems (Sousa et al., 2017). Among these systems, platforms offering pre-recorded lectures have gained significant attention due to their ability to provide flexible content anytime to large audiences (Hadgu et al., 2016). Despite their widespread use, however, traditional pre-recorded lecture systems are constrained by several factors. First, they offer static content that typically comes in the form of videos or slides, which cannot meet the diverse learning needs and preferences of students. Second, these systems often fail to adapt to different learning styles or facilitate meaningful interactions between teachers and students. This lack of interactivity and personalization often dulls the learning experience and reduces information retention.

The abovementioned issues have been addressed through solutions such as embedded quizzes, scripted chatbots, and recommendation algorithms, but these measures often fall short in the delivery of instantaneous guidance, emotional resonance, and adaptability to individual learning needs. These limitations highlight the need for a more holistic and responsive approach to online education. In response to this call, we developed an interactive, artificial intelligence (AI)-driven pre-recorded lecture system

Received January 22, 2025; revised June 11, 2025; accepted July 7, 2025

Kai Zhang (✉)

E-mail: kkzhang08@ustc.edu.cn

that optimizes traditional content delivery. Our system combines a range of innovations designed to create a more engaging, personalized, and adaptive learning environment (see [Figure 1](#)). The left panel depicts conventional materials (PDF files and PowerPoint presentations), which lead to one-way student learning and a lack of real-time feedback. The right panel indicates the pre-recorded lecture, which features two-way interaction between students and the digital teacher, resulting in personalized learning experiences.

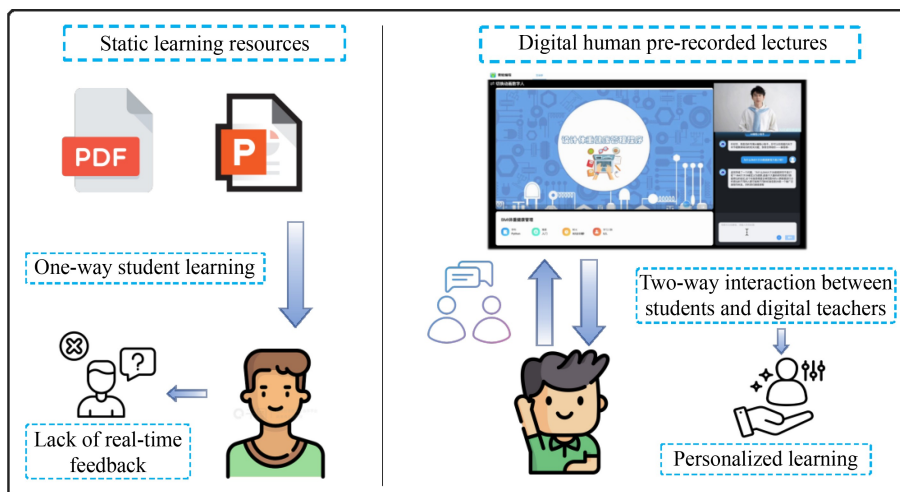
Digital human technology is key to our system. The use of digital humans in educational settings presents multiple advantages, as these avatars can effectively engage students through personalized interactive experiences. By replicating behaviors such as emotional expression and body language, these virtual agents can simulate the presence of live teachers, which is especially important in virtual education, where physical absence can cause learners to lose interest. Digital humans can also help address issues of educational inequality by providing excellent teaching to areas with limited access to human instructional resources. Moreover, their scalability ensures the delivery of individualized education to numerous students without compromising the quality of the learning experience. Finally, using digital humans can streamline content delivery, with virtual figures managing content distribution tasks, while educators focus on higher-order responsibilities, such as guidance provision and content curation.

Notwithstanding the recent increase in the sophistication of digital teaching technology, current instructional delivery via digital human teachers still faces significant challenges, among which a frequently occurring problem is their lack of insight into learners' needs. This is especially true in open-ended dialogues, during which AI-generated instructors cannot

determine user intentions and learning preferences from conversations, leading to discrepancies in their understanding of learners' cognitive states. Difficulties are also encountered in aligning learners' knowledge levels with the capabilities of existing large language models (LLMs). Due to the incomplete development of the personalized long-term memory functions of digital human teachers, these agents cannot track students' learning trajectories or summarize learning preferences, which limits the effectiveness of the so-called personalized instruction offered through these technologies.

However, thanks to the rapid development of LLMs ([Naveed et al., 2023](#); [Zhao et al., 2023](#)), the textual comprehension abilities of digital human teachers have considerably matured. Agents driven by LLMs are considered a cornerstone of the technology behind digital human teachers, but these models remain encumbered with challenges in practical application, such as constrained understanding due to unimodal communication ([Jiang et al., 2025](#)) and insufficient depth in addressing personalized needs despite their broad generalization capabilities ([Eapen & Adhithyan, 2023](#)). The absence of robust verification mechanisms can also result in automatically generated knowledge content that occasionally includes inaccurate information, which is a significant drawback of these innovations. Additionally, delays in updating knowledge bases may introduce outdated or erroneous information into generated content. These problems highlight the criticality of ensuring the accuracy and reliability of knowledge content for virtual or digital platforms.

The challenges confronting virtual human technology also relate to aspects of interactive naturalness, especially emotional recognition and contextual understanding. Compared with real teachers, digital humans lack emotional perception and



**Figure 1** Contrast between traditional static learning resources and pre-recorded lectures delivered by digital humans.

expression, thus giving rise to difficulties in the provision of emotional support and empathetic care during interactions. Finally, current virtual individuals often undergo context switching and deliver contradictory responses when handling continuous conversations, which generates unnaturally flowing dialogues that diminish user immersion and learning satisfaction.

Against this backdrop, we designed a comprehensive digital human teaching system that overcomes the aforementioned difficulties while meeting educational needs. Our system employs various advanced AI technologies to assist in generating recorded lessons to be delivered by digital human instructors. Specifically, the system uses an LLM whose capabilities are fine-tuned on the basis of educational materials (e.g., PDF files and PowerPoint presentations) to improve the comprehension of content while generating coherent teaching materials relevant to context. It then leverages AI and computer vision technologies to create personalized digital lecturers whose virtual representations mimic human features, such as facial expressions, lip-syncing, and body language, thus providing an engaging teaching presence that resonates emotionally with students. The system is also integrated with an automatic content generation process that seamlessly combines personalized instructors with educational materials and creates excellent lesson videos aligned with contextual requirements. Additionally, our system includes an interactive question-and-answer (Q&A) feature that allows students to ask questions after watching recorded lessons. The system uses data from student profiles to deliver tailored, empathetic, and context-sensitive responses, thereby cultivating a dynamic education environment and further enhancing the learning experience.

By integrating these innovations, we aim to bridge the gap between static pre-recorded content and dynamic real-time interaction. This not only improves content delivery but also addresses key issues of personalization and interactivity, providing students with a more engaging and effective learning experience. Overall, the proposed system represents a significant advancement in recording technology for courses, offering a more immersive, scalable, and personalized educational process that satisfies the evolving needs of students and educators.

## 2 Related Works

### 2.1 | Large Language Models in Education

The application of LLMs in education has experienced rapid growth in recent years, opening up new

possibilities for enhancing learning experiences (Kasneji et al., 2023; Yan et al., 2024). LLMs based on GPT architectures (Achiam et al., 2023; Floridi & Chiriatti, 2020) have shown remarkable potential with respect to understanding human responses and generating human-like text, making them valuable tools for tasks such as content generation (Rashid et al., 2024), student interaction (Kumar et al., 2023), and personalized learning (Razafinirina et al., 2024).

A primary application of LLMs in education is the automated generation of teaching materials (Hu et al., 2024). These models can process large volumes of educational content, extract key information, and convert it into structured instructional resources that include summaries, practice questions, and explanations. For example, they can produce customized lesson plans or adapt content to suit various proficiency levels (Karpouzis et al., 2024), ensuring accessibility for a diverse learner population. This capability helps address the resource-intensive nature of manual content creation and allows for rapid adaptation to evolving curricula.

LLMs are also instrumental in the provision of interactive tutoring and support (Lieb & Goel, 2024). As virtual tutors, they provide immediate feedback, respond to student inquiries, and even simulate Socratic questioning that fosters critical thinking (Izumi et al., 2024). Their advanced natural language processing abilities enable them to understand complex queries and engage in human-like dialogue, creating an interactive and immersive learning experience. LLMs likewise contribute to personalized learning through student data analysis meant to identify strengths, weaknesses, and preferences (Razafinirina et al., 2024). This analysis enables the development of personalized recommendations and adaptive learning pathways, which have been shown to enhance learning outcomes.

Despite the advantages of LLMs, however, their integration into education presents challenges such as the potential lack of domain-specific knowledge among general-purpose models (Welz & Lanquillon, 2024). While these technologies excel at generating coherent responses, they may produce inaccurate or irrelevant information in specialized fields without proper fine-tuning. The refinement of LLMs on the basis of educational datasets specific to a given domain is an ongoing area of research aimed at improving the accuracy and relevance of their outputs. Attention must also be paid to addressing ethical issues, including bias, data privacy, and the risk of overreliance on automated systems (Jiao et al., 2024; Liyanage & Ranaweera, 2023), to guarantee responsible and effective integration into educational contexts.

### 2.2 | Digital Human Technologies

Digital human technologies have advanced considerably, with such innovations producing avatars

that replicate human appearance, voice, and behavior (Burden & Savin-Baden, 2019; Sun et al., 2023; Zhang et al., 2025; Zhen et al., 2023). These technologies are widely used in gaming, entertainment, customer service, and education, where they enhance empathy and engagement during human-machine interactions (Kimani et al., 2021; Paiva et al., 2017). Digital humans are integrated with technologies such as computer vision, natural language processing, speech synthesis, and motion capture. For example, generative models create realistic facial expressions and speech, while neural text-to-speech (TTS) synthesis mimics human tone and rhythm (Ren et al., 2019).

These capabilities are ideal for applications requiring empathetic communication, such as virtual teaching. A key breakthrough is the use of transformer-based models (Chopin et al., 2023), which enable dynamic expressions and gestures that simulate emotion and energy. These features enhance educational experiences in settings where an instructor’s expressions are critical to learning. Personalization is also a focus, with avatars customizable to represent specific educators and adapt to learner needs (Fiedler et al., 2023; Wolf et al., 2022). This customization is achieved through style transfer techniques in both visual and auditory domains. As can be seen, digital humans can be used to create interactive learning environments, with these agents serving as tutors or conversational partners (Ieronutti & Chittaro, 2007). Despite their potential, however, challenges remain, including the computational demands for real-time interaction and the uncanny valley effect, problems that require ongoing improvements in rendering and user experience design.

### 2.3 | Pre-Recorded Lectures

Pre-recorded lecture technologies have become a cornerstone of digital education, paving the way for flexibility and accessibility in asynchronous learning environments (Le, 2022). These systems allow students to learn at their own pace and revisit content as necessary. Early formats primarily involved video

recordings or narrated slides, but although these are simple to implement, they often lack features that promote active learning, such as personalized feedback or real-time interactivity (Kim & Kim, 2024). This limitation can diminish student motivation and retention relative to that achieved during live instruction.

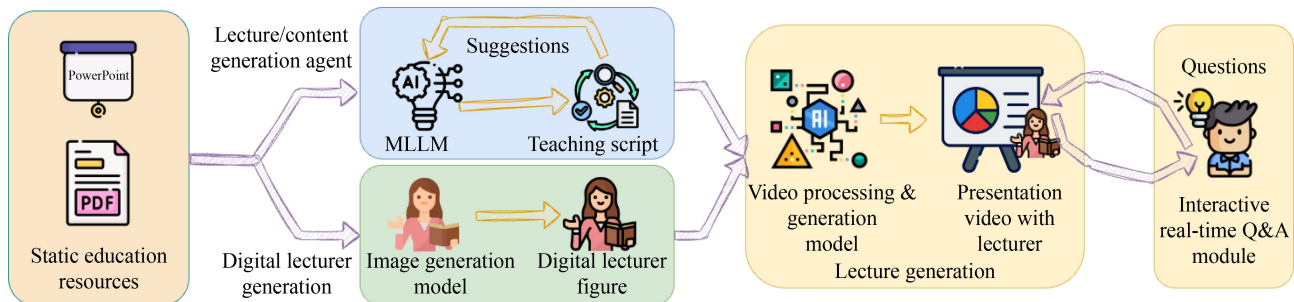
Recent advancements have been directed toward overcoming these challenges by integrating interactive components into pre-recorded lectures (Secker et al., 2022). Features such as embedded quizzes, annotations, clickable resources, and branching scenarios enable students to assess their understanding, explore supplementary content, and choose their own learning paths, which increase their engagement and participation. AI-driven analytics is also pivotal in enhancing the learning experience. These systems track student interactions (Liu et al., 2021), identify areas where learners struggle, and generate adaptive content to address common issues, thus personalizing the education process (Sheng et al., 2023). Nevertheless, many existing pre-recorded lectures continue to be static in nature, restricting immersion and interactivity.

## 3 Our Framework

The integration of digital human technology into our proposed system offers a potential solution to the challenges discussed earlier, as this provides a more dynamic and interactive learning environment.

### 3.1 | Pipeline

The pipeline of our interactive digital instructor-based pre-recorded lecture system integrates various advanced AI technologies, thereby ensuring seamless flow from resource input to personalized content delivery and interactive learning. The system follows a structured process that ensures exceptional, contextually relevant, and adaptive educational experiences (see Figure 2).



**Figure 2** Interactive digital instructor system: End-to-end pipeline. MLLM: multimodal large language model, Q&A: question-and-answer.

This pipeline illustrates the seamless transformation of educational resources into engaging digital lectures. From resource input and AI-driven content analysis to digital instructor customization, video generation, and interactive Q&A modules, the system ensures personalized, adaptive, and high-quality learning experiences for modern education needs.

At the core of the pipeline are the extraction and understanding of educational resources, followed by the creation of a personalized digital instructor and the generation of synchronized lecture videos. The first step in the pipeline involves the ingestion of educational resources, such as PDF files or PowerPoint presentations, which are processed by a fine-tuned LLM. This model can analyze textual content, identify key concepts, and generate a coherent and pedagogically appropriate teaching script. The output is a structured collection of content, ready to be delivered in an appealing format.

Once content has been understood and structured, the system moves on to the second stage, the customization of the digital human instructor. Using advanced computer vision and synthesis techniques, the system creates a personalized teacher avatar on the basis of an image of a real instructor. This avatar is designed not only to resemble the instructor visually but also to replicate nonverbal cues, such as lip-syncing, facial expressions, and bodily gestures, to ensure a lifelike and compelling teaching presence. This step is essential for maintaining a strong connection between the instructor and the learner, simulating the dynamic nature of in-person teaching.

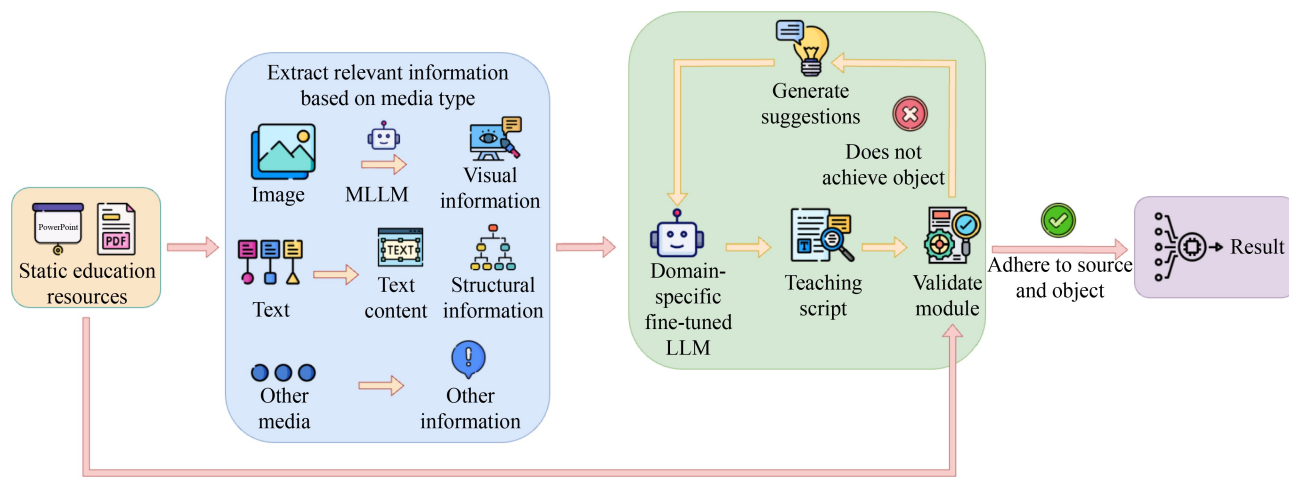
The third stage of the pipeline is the automated generation of a lecture video, in which the system combines a teaching script with the personalized instructor, synchronizing speech, gestures, and visual elements to produce outstanding video material. This

video is intended to be as interactive and engaging as a live lecture, incorporating not only spoken content but also visual aids that include diagrams, charts, and slides to enhance understanding.

Finally, the system comes with an interactive Q&A module, allowing students to ask personalized questions after a lecture. The system analyzes a student’s user profile and learning history to provide tailored, context-aware, and empathetic responses. This stage ensures that the system remains responsive to individual learning needs, fostering an interactive and adaptive learning experience. The pipeline is meant to operate in a harmonious manner, with each stage building on the previous one to create a cohesive, personalized, and engaging educational experience. In the subsequent sections, we explore the specifics of each stage, including the underlying technologies and models that enable the system to function efficiently and effectively.

### 3.2 | Resource Understanding

The effective transformation of static educational resources, such as PDF files and PowerPoint slides, into dynamic teaching content requires an in-depth grasp of these materials. At the heart of our system’s resource understanding module lies a fine-tuned LLM that serves as the foundation for analyzing, organizing, and contextualizing input resources. While general-purpose LLMs have demonstrated impressive language understanding and generation capabilities, they often lack the domain-specific knowledge and pedagogical alignment necessary to create high-quality educational content. These limitations are addressed by our system through a domain-specific fine-tuning approach (see Figure 3).



**Figure 3** The resource understanding module transforms static educational resources into dynamic, pedagogically aligned teaching scripts using domain-specific fine-tuned LLMs and structured content analysis. LLM: large language model; MLLM: multimodal LLM.

A base LLM is adapted using a curated corpus of educational materials that not only represent the target domain but also encompass teaching strategies and best practices. This refinement process enables the model to grasp the nuances of the terminology, instructional styles, and content organization that typify a given domain. The fine-tuned model can therefore generate accurate and pedagogically effective teaching scripts.

The resource understanding process begins with content absorption, wherein the system parses input files to extract textual and structural information. This includes identifying key concepts, hierarchies, and relationships within the materials. In a PowerPoint presentation, for instance, the model can distinguish between slide titles, bullet points, and supplementary notes, ensuring that each element is appropriately contextualized in the generated script. Similarly, in a PDF document, the system recognizes headings, subheadings, and inline annotations, after which it reconstructs the logical flow of the content.

The fine-tuned LLM is essential to the transformation of extracted raw information into coherent teaching narratives. Leveraging its contextual understanding functionality, the model organizes content into a logical sequence, providing clear explanations, examples, and transitions. It also identifies opportunities to integrate additional pedagogical elements, such as analogies, questions, or summaries, thereby enhancing student engagement and comprehension. Furthermore, our system is equipped with mechanisms designed to guarantee the reliability and relevance of generated content. For domains

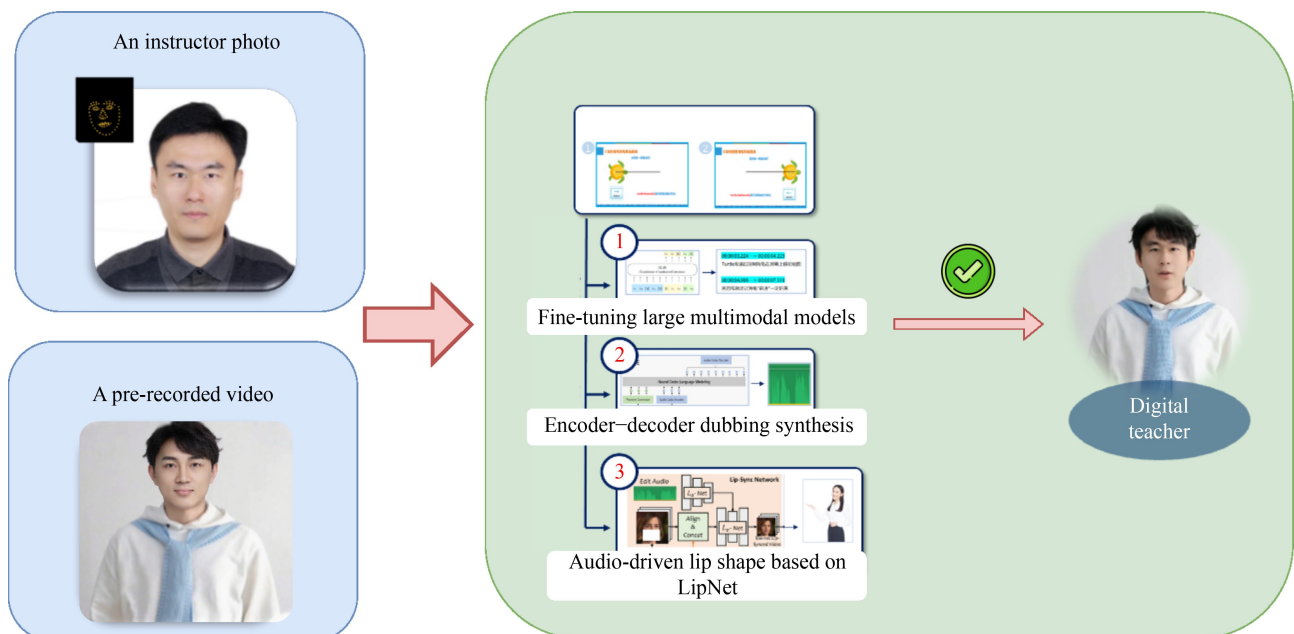
characterized by rapidly evolving knowledge, such as science or technology, the system can be periodically updated with new domain-specific data to maintain its accuracy and relevance. Additionally, the fine-tuned LLM comes with validation layers that compare generated output with input resources, ensuring fidelity to source materials and adherence to pedagogical objectives.

To sum up, the resource understanding module combines the power of fine-tuned LLMs with tailored preprocessing and postprocessing strategies, which enables it to convert static resources into dynamic, pedagogically sound teaching scripts. By addressing the challenges associated with domain knowledge gaps and contextual relevance, our approach lays a solid foundation for the succeeding stages of digital instructor customization and lecture video generation.

### 3.3 | Generation of a Personalized Digital Instructor

The creation of a personalized digital instructor is a critical component of our framework, with state-of-the-art techniques in computer vision, audio synthesis, and generative modeling put to advantage in this process. This module is aimed at producing a lifelike digital representation of a human instructor that faithfully reproduces their visual appearance, voice characteristics, and expressive traits while ensuring alignment with teaching content (see Figure 4).

The personalized digital instructor generation module creates a highly realistic and personalized digital instructor by integrating advanced computer



**Figure 4** The process of digital instructor generation.

vision, audio synthesis, and generative modeling, ensuring the seamless synchronization of visual, auditory, and motion features to deliver engaging, interactive teaching experiences that closely mirror the original instructor's appearance, voice, and expressive traits.

The process begins with the preprocessing of an input video of an instructor, which provides the foundational data for constructing a digital avatar. This step entails extracting synchronized image, video, audio, and textual data, ensuring clean and structured input for subsequent stages. To capture the auditory identity of the instructor, the system processes audio data through a pipeline dedicated to speech synthesis. The audio material is analyzed to derive two key features: style and phonemes. The style encoder identifies unique voice characteristics, such as pitch, tone, and speaking style, while the phoneme encoder translates the accompanying textual data into phonetic sequences. These sequences are then adjusted using a variance adapter to match natural speech rhythms. Together, these components clear the way for the system to construct natural and instructor-specific speech audio.

Simultaneously, the visual and motion characteristics of the instructor are extracted from the uploaded video. The system employs an image encoder to identify facial features, capturing the instructor's appearance and expressions. A motion encoder analyzes dynamic features, including gestures and head movements, guaranteeing that the generated digital instructor reflects realistic motion patterns. An audio-motion synchronization module integrates the generated audio features with motion characteristics to produce coherent, synchronized gestures and lip movements.

The final step in the generation process involves video synthesis. Using the extracted visual and motion features, as well as the synthesized audio material, a video decoder produces a personalized, high-quality video of the digital instructor. Advanced techniques, such as noise augmentation and denoising, are employed to enhance the robustness and realism of the output. The synchronization of audio and video materials ensures that lip movements and expressions align perfectly with spoken content, creating a seamless and dynamic visual experience.

To optimize the generation model, training is conducted with a combination of reconstruction and variance losses. The use of reconstruction loss ensures that the generated output closely resembles the original instructor in both appearance and auditory fidelity, while the adoption of variance loss improves the temporal consistency of synthesized motions. This approach represents a significant advancement in the creation of personalized digital instructors. By

combining multimodal data processing with cutting-edge synthesis techniques, the system achieves substantial realism and personalization, enhancing a learner's engagement and connection with content. This module thus serves as a pillar of our framework, bridging the gap between traditional static content delivery and dynamic, interactive teaching experiences.

### 3.4 | Video Content Generation

The system proceeds to the synthesis of excellent educational videos, during which multimodal data, including audio, visual, and motion features, are integrated to create compelling and cohesive lecture videos that replicate the dynamism of in-person teaching. Video generation commences with the synchronization of the teaching script and the personalized instructor's speech and gestures. The teaching script is converted into speech using a high-fidelity TTS system, which is designed specifically to match the tone, rhythm, and style of the instructor's original voice modeled during the instructor generation phase. The resulting audio material serves as the temporal backbone for the video output, ensuring that all subsequent elements align perfectly with the speech.

Simultaneously, the motion and facial features of the digital instructor are animated for synchronization with the generated audio. The system uses a speech-driven animation framework, wherein phoneme-level timing from the TTS output guides the instructor's lip movements, while prosodic cues, such as intonation and stress, inform head movements and facial expressions. This guarantees that the digital instructor's gestures and expressions are contextually appropriate and enhance content delivery. The visual component of the video is rendered using the digital instructor's personalized appearance, which was modeled in the previous stage. This process encompasses the accurate reproduction of facial details, body language, and clothing. To maintain engagement, the system dynamically integrates supplementary materials, such as slides, diagrams, or text highlights, alongside the instructor video. These visual aids are presented synchronically with the instructor's speech, reinforcing key concepts and aiding comprehension.

To advance smooth and natural flow in the presentation, the system makes use of advanced temporal alignment mechanisms intended to meticulously fine-tune transitions among speech, gestures, and visual aids. These mechanisms harmonize the timing and coordination of various elements, ensuring that each component seamlessly complements the others. For example, the pace of speech delivery is dynamically adjusted on the basis of the complexity of content, allowing learners to follow along more effectively without feeling rushed or overwhelmed.

Simultaneously, the system maintains consistent visual focus on the most relevant elements, such as key visuals or gestures, guaranteeing that audience attention is directed toward the most critical aspects of the presentation.

The final output is a polished educational video that combines the expressiveness of the personalized digital instructor with the clarity of well-structured teaching content. By closely mirroring the dynamics of live instruction, the system not only delivers content effectively but also fosters a more appealing and immersive learning experience for students. This stage completes the automated pipeline, and the resulting video serves as the foundation for the interactive learning experiences described in the next module.

### 3.5 | Interactive Question-and-Answer System

Conventional pre-recorded lectures typically offer students little to no opportunity for interaction. To resolve this constraint, we equipped our system with an interactive Q&A module, harnessing the strengths of digital content delivery while addressing its primary drawbacks. This module enables real-time interaction as well as incorporates advanced personalization and empathetic engagement into the teaching process, thereby enhancing the overall learning experience. At the core of the Q&A system is a fine-tuned LLM capable of understanding and responding to a wide range of student queries. Following the completion of a lecture, students can pose questions, which the system processes by extracting their contextual and semantic meanings to generate precise, relevant, and pedagogically sound responses. This functionality provides students with a sense of immediate assistance, mirroring the benefits of live instruction.

To further elevate the Q&A experience, the system is integrated with personalized services grounded in the diagnostic assessment of a student. Using input from the student's interaction history, performance metrics, and stated goals, the system constructs a dynamic user profile that captures their learning preferences, strengths, and challenges. This personalized profile enables the Q&A module to tailor its responses to the student's particular needs. For example, the system provides simplified explanations and additional resources to students who struggle with foundational concepts, but it offers deeper insights and challenges to advanced learners.

Empathy is strategically integrated into the system's architecture through a multilayered empathy model, bringing forth a paradigm shift from traditional emotion-agnostic educational AI. The system employs an empathy-driven natural language processing framework that combines (1) real-time sentiment

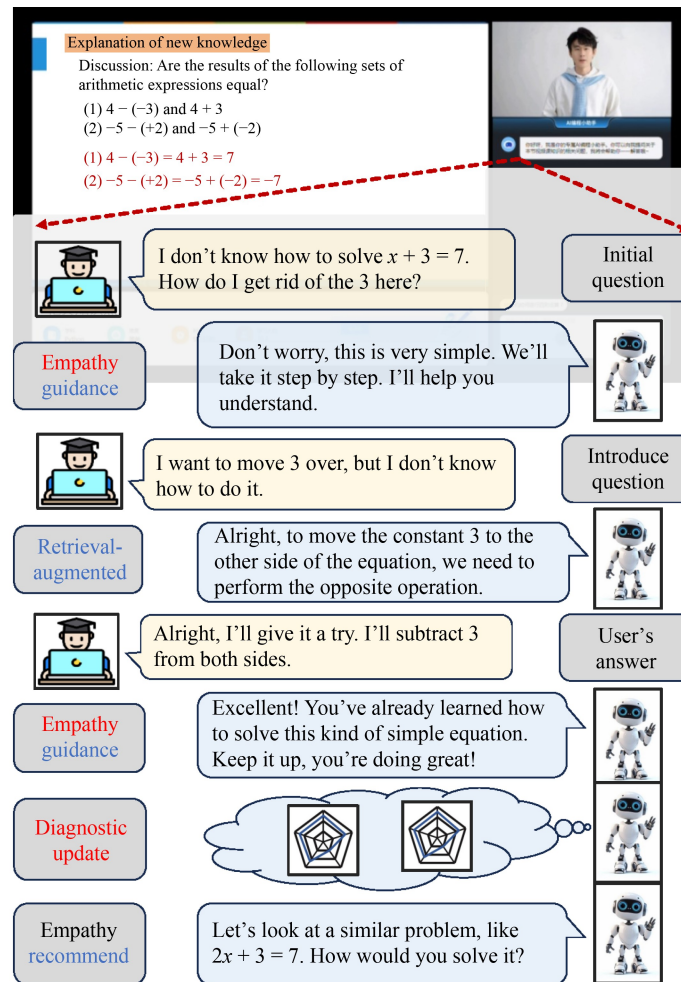
analysis using transformer-based emotion classifiers (Sun et al., 2024; Zhang et al., 2023), (2) context-aware emotional resonance algorithms, and (3) pedagogically aligned response generation with affective scaffolding. This tripartite approach enables the digital instructor to both detect and authentically react to learners' emotional states through what we term cognitive-affective mirroring, a technique that adapts both content delivery and emotional tone on the grounds of detected psychological states.

The empathy model operates through a dual-channel processing system: While the cognitive channel analyzes semantic content for knowledge gaps, the affective channel evaluates lexical choices, syntactic patterns, and temporal response characteristics to infer emotional states (Sun et al., 2024). The case study illustrated in Figure 5 shows a Q&A process, which reveals the effectiveness of the system in delivering personalized learning and motivating student interest in learning. By implementing a dynamic empathy calibration mechanism that adjusts support strategies on the basis of both immediate emotional cues and longitudinal learning patterns, the system achieves what we characterize as adaptive pedagogical attunement, maintaining optimal challenge levels while preserving learner self-efficacy. The module transforms static lectures into dynamic, student-centered experiences. It tailors responses on the basis of user profiles, fosters motivation through emotional intelligence, and adapts content delivery to address individual learning needs for deeper understanding and growth.

The interactive Q&A module also supports adaptive content delivery. With insights derived from a student's questions and profile as bases, the system can recommend supplementary materials, suggest relevant topics for further exploration, or even modify future lesson plans to address identified gaps. This feedback loop ensures that the learning experience remains dynamic and responsive to the student's evolving needs. By combining real-time interaction, personalized diagnostics, and empathetic engagement, our Q&A system transforms the static nature of traditional pre-recorded lectures into a dynamic and student-centered learning environment. This module not only reconciles asynchronous content delivery and live instruction but also maximizes the advantages of digital teaching, advancing exhaustive understanding, sustained interest, and personalized growth for every learner.

## 4 Case Study: Resource Understanding

Since our research primarily focused on system design and implementation rather than purely theoretical analysis, we conducted case studies to illustrate the



**Figure 5** The interactive Q&A module enhances digital learning by combining real-time interaction, personalized diagnostics, and empathetic engagement.

practical effectiveness of our approach. By demonstrating how our system transforms static educational resources into dynamic, interactive teaching scripts, we intended to provide qualitative insights into its pedagogical impact.

This case study examined the effectiveness of a fine-tuned LLM in transforming static educational resources into dynamic, interactive teaching scripts. We compared its performance to that of a general-purpose LLM, Llama3, in terms of content accuracy and pedagogical quality. The Llama3 model was used in its default configuration without fine-tuning and prompted with the same educational materials as those used for our fine-tuned model. Prompts were designed to elicit explanatory teaching content, and the generated scripts were collected directly for evaluation. This setup allowed for a fair comparison of the teaching clarity and instructional quality provided by the general and domain-specific educational LLMs. We posited that our fine-tuned model would outperform the general LLM, yielding better results in both teaching efficacy and student engagement.

## 4.1 | Setup and Implementation

A total of 67 participants (60 students and 7 teachers) from the computer science department of a university participated in this study. The educational materials consisted of three topics from an introductory course on data structures and algorithms. The course materials covered fundamental topics, including array manipulation, linked list traversal, and basic sorting techniques. Both the fine-tuned and general-purpose LLMs were tasked with generating teaching scripts with these materials as references.

The students were randomly assigned to one of three groups. Group A was exposed to the teaching scripts generated by the fine-tuned LLM and specifically tailored to the course content. Group B received teaching scripts produced by the general-purpose LLM, with the content uncustomized to the particular subject. Group C, the control group, was provided with conventional educational resources, such as PDF files and PowerPoint slides, without dynamic teaching scripts or content generated by any LLM.

## 4.2 | Evaluation

Each student participated in a 30-minute learning session, after which all the participants were administered a quiz to assess their comprehension of the material. A survey was also conducted among the teachers to gather feedback on the perceived clarity and relevance of the content. Both the students and teachers provided insights regarding the quality of the teaching scripts (Groups A and B) and the effectiveness of the static resources (Group C).

## 4.3 | Results

The results demonstrated a clear advantage of the fine-tuned LLM. Group A, which interacted with this model's scripts, scored an average of 90 on the quiz, outperforming both Group B (83, general LLM) and Group C (80, control). This superior performance can be attributed to the tailored pedagogical approach underlying our model. It incorporated domain-specific examples, analogies, and a logical flow of content into the session, which advanced the comprehension and retention of the material.

Out of the seven teachers, six rated the scripts generated by the fine-tuned LLM as clear, accurate, and pedagogically effective. They particularly appreciated the integration of domain-specific terminology and instructional strategies, which enhanced student engagement and understanding. Five teachers expressed concern that the general LLM-produced scripts lacked sufficient depth and instructional techniques. While the content was accurate, it did not engage students as effectively, missing the pedagogical elements needed for profound learning. All the teachers noted that while the static resources (PDF files and slides) were useful, they failed to stimulate the level of engagement and support necessary for extensive understanding, especially when complex concepts were tackled.

## 4.4 | Discussion

The findings suggest that fine-tuning an LLM with domain-specific educational content significantly enhances its ability to generate dynamic, engaging, and pedagogically sound teaching scripts. Compared with the general-purpose model, the fine-tuned LLM demonstrated superior content accuracy and teaching effectiveness, which ultimately enriched student learning outcomes and engagement levels.

This study underscores the potential of our fine-tuned LLM for use in educational contexts, where it can supplement traditional teaching materials and offer personalized, contextually relevant instruction. Future work will focus on polishing the fine-tuning process to further optimize the system's performance

and expand its application across different educational domains.

# 5 Case Study: Programming Education

To assess the effectiveness of the proposed digital human teaching system in enhancing both engagement and learning outcomes, a controlled A/B testing experiment was conducted with a group of primary school students receiving children's programming education. The study compared two distinct teaching modalities: one using a traditional teacher-led video and the other employing a digital human as the instructor. The study assessed how the system's personalized digital teacher and interactive Q&A features can improve young learners' engagement, comprehension, and creativity in programming tasks.

## 5.1 | Setup and Implementation

The participants were primary school students aged 8 to 12 years who were enrolled in an introductory programming course. The instructional content focused on foundational Scratch programming concepts, including sequence logic, loops, and simple event-driven design. The experiment was conducted over a period of four weeks with two groups of students: Group A received traditional recorded lectures, while Group B engaged with a digital human avatar as their instructor. The digital human used advanced technologies to replicate human-like interaction, providing verbal explanations, visual cues, and empathetic feedback to enhance student engagement.

## 5.2 | Evaluation

Both groups completed the same set of programming exercises and were assessed with respect to their learning outcomes through a series of quizzes. They were also administered surveys to evaluate their motivation and engagement levels throughout the course. Specifically, engagement was measured using a questionnaire that included items such as the following:

- "How interesting did you find the lessons?"
- "How motivated were you to complete the exercises?"
- "Did you feel that the teacher helped you when you had difficulties?"

Learning outcomes were determined using a posttest designed to assess the knowledge retention and programming skills of the students. The test consisted of multiple-choice questions and coding challenges based on the lessons taught during the course. The experiment also incorporated A/B testing, wherein Group A's performance served as the baseline against

which Group B's performance was measured to determine significant differences in outcomes due to the digital human instruction.

### 5.3 | Results

Group B, which interacted with the digital human avatar, reported significantly higher engagement scores than those achieved in Group A. Specifically, 85% of the students in Group B described the lessons as "interesting" or "very interesting" compared with 65% in Group A. The students in Group B were also more likely to acknowledge the avatar's assistance in overcoming learning obstacles, with 75% of them agreeing with this statement compared with only 55% in Group A.

In terms of learning outcomes, Group B performed better on the posttest, earning an average score of 87, whereas Group A's average score was 75. The improvement in Group B was statistically significant ( $p < 0.05$ ), suggesting that the digital human avatar contributed positively to knowledge retention and skill acquisition.

### 5.4 | Discussion

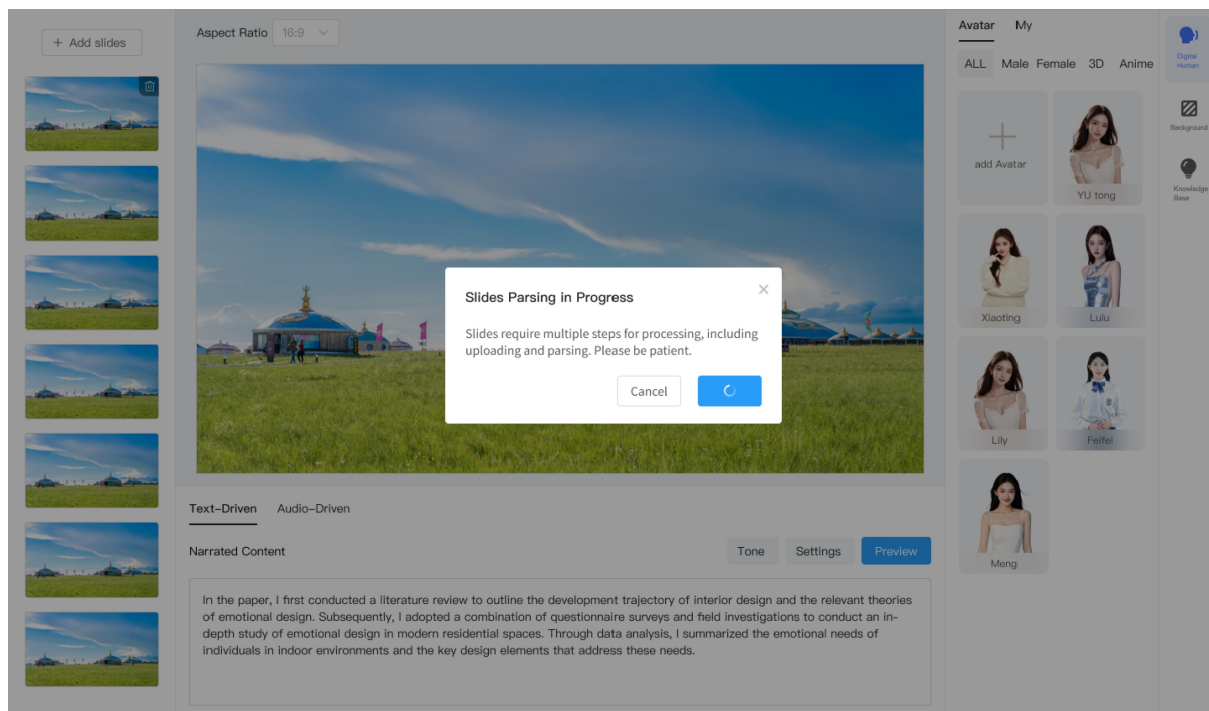
This case study demonstrated that the use of digital human avatars in a children's programming course significantly enhances both student engagement and learning outcomes. The higher engagement levels in

Group B suggest that the empathetic and interactive nature of the digital human instructor fostered a more motivating and enjoyable learning environment. This finding aligns with previous research showing that digital humans can increase engagement and empathy in educational settings.

The higher posttest scores in Group B imply that the digital human's ability to provide personalized feedback and adjust to individual learning needs contributed to improved knowledge retention. However, while the results are promising, it is important to consider potential limitations, such as the computational resources required to run such systems and the possible challenges in designing avatars that can effectively mimic human interactions without triggering the uncanny valley effect. Overall, this study highlights the potential of digital humans to transform the landscape of education, providing a scalable, interactive, and personalized learning experience for children.

## 6 System Demonstration

The system demonstration foregrounds advancements in digital humans for recorded courses, showcasing their ability to enrich learning through personalized interaction. The process begins with the system's ability to upload and process static PowerPoint content (see [Figure 6](#)), in which key elements such as texts,



**Figure 6** Uploading and understanding of PowerPoint content. This figure focuses on the resource-processing mechanism that underpins downstream learning interactions.

structures, and visuals are extracted using a fine-tuned, domain-specific LLM. This enables the transformation of static resources into dynamic, pedagogically effective teaching scripts tailored to diverse educational needs. Building on these features, the system introduces personalized digital teacher avatars (see Figure 7), whose visual and interactive characteristics are adapted to accord with individual learners' preferences. These avatars foster relatability and engagement, ensuring a more personalized and effective teaching style. Finally, the system delivers a student-centered learning experience (see Figure 8) by enabling real-time interaction, personalized diagnostics, and empathetic responses. By addressing knowledge gaps, offering tailored suggestions, and enhancing understanding, the system showcases the potential of digital humans to create immersive, engaging, and highly effective learning environments.

## 7 Future Outlook

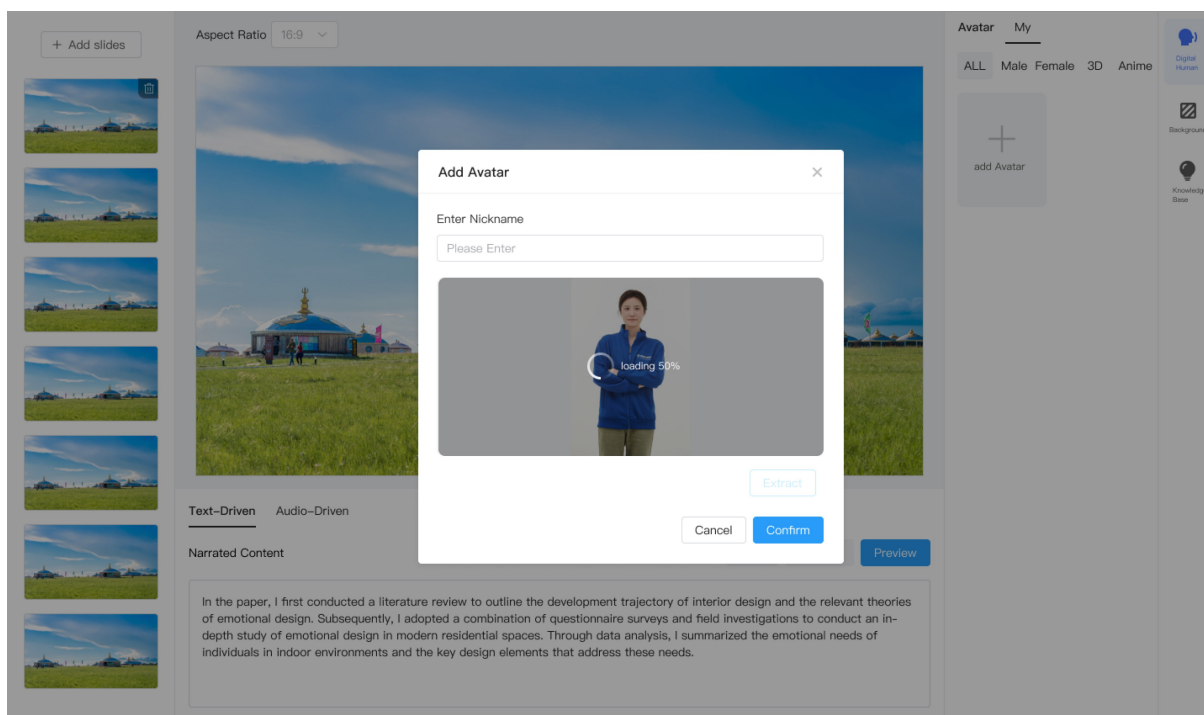
Our proposed system reflects the promise of integrating LLMs, personalized digital teacher technology, and pre-recorded lecture systems in improving educational outcomes. While this study centered on system design and feasibility, future work will involve more extensive empirical validation, including controlled experiments and large-scale user studies, to further substantiate its

impact on learning outcomes. Several other areas remain open for exploration to advance educational technologies in this domain. These initiatives are aimed at expanding the system's applicability, optimize its educational impact, and address potential technical and ethical challenges.

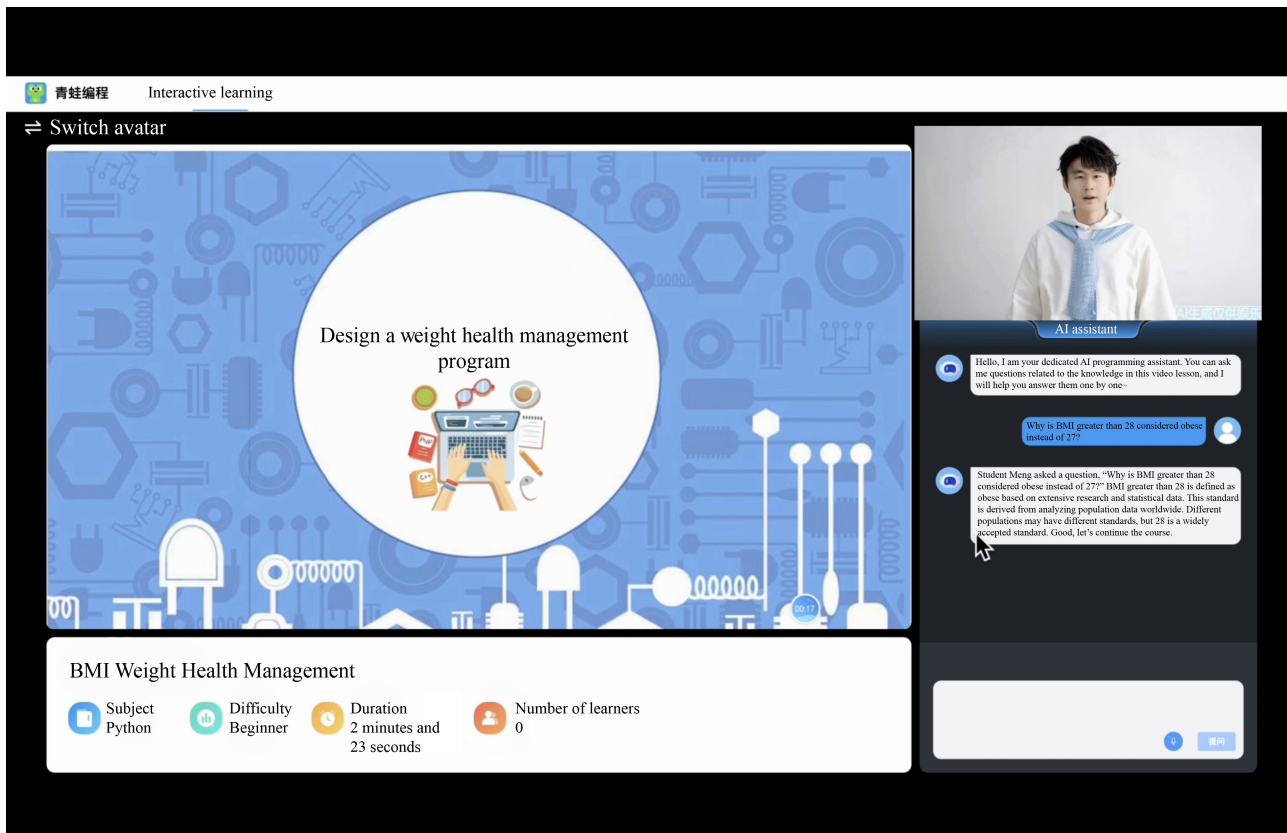
### 7.1 | Broader Educational Applications

While the current system primarily emphasizes real-time feedback and short-term interactions, future iterations can incorporate longitudinal learning analytics to monitor learners' progress over time. By leveraging large-scale data analysis, the system can pinpoint knowledge gaps, evaluate the depth of learners' understanding, and monitor skill development. These features would facilitate the creation of personalized, long-term learning plans for students while offering valuable insights that can advance enriched curriculum design.

The system's potential extends vastly beyond formal education settings. It can be tailored to vocational training, producing customized digital instructors for specific industries or job roles to accelerate the acquisition of practical skills. It can also support lifelong learning by generating interdisciplinary content aligned with users' interests and goals, as is the case with courses that blend science, art, and history. In underserved regions, including rural areas and



**Figure 7** Personalized digital teacher avatars, with their visual and interactive characteristics dynamically adapted to align with individual learners' preferences, fostering engagement and relatability. The figure also illustrates the avatar customization logic behind instructional content generation.



**Figure 8** Provision of a dynamic, student-centered learning experience through real-time interaction, personalized diagnostics, and empathetic responses, which addresses knowledge gaps, offers tailored suggestions, and enhances learners' engagement and understanding. This interface represents the front-end experience, wherein learners interact with the digital teacher.

community education spaces, the system can deliver outstanding, interactive content to populations with limited access to traditional educational resources, thereby helping reduce educational disparities.

Furthermore, teaching plans across different disciplines may require distinct instructional strategies and content presentation formats. For example, humanities subjects, such as literature or history, often emphasize open-ended discussions, narrative analysis, and critical thinking, which would benefit from conversational dialogue and contextual storytelling via a digital instructor. In contrast, Science, Technology, Engineering, and Mathematics (STEM) subjects typically demand structured explanations, visual simulations, and procedural demonstrations. Our system framework is designed with modularity in mind, allowing for the future customization of digital human behaviors, content generation styles, and interaction patterns for improved correspondence with the pedagogical characteristics of each field.

## 7.2 | Ethical and Psychological Considerations

As digital teachers become increasingly advanced,

ethical and psychological considerations must be addressed. Principal concerns include the degree to which learners trust digital teachers compared with human instructors and whether prolonged reliance on these systems can undermine students' independent learning skills. Future research can explore how to incorporate transparency and user control into system design. For instance, providing options for learners to adjust the level of personalization or the expressive style of their digital teacher may improve user comfort and autonomy. Privacy concerns should also be prioritized, with robust data protection mechanisms implemented to ensure the security of learners' personal information and maintain user trust.

## 7.3 | Scalability and Accessibility

To maximize the system's impact, future developments should revolve around scalability and accessibility. Lightweight versions optimized for low-power devices would render the system accessible in regions with deficient technological resources. A flexible customization framework would be crucial to accommodate diverse curricula and teaching methodologies in education systems across the world.

Moreover, the development of open-source platforms is anticipated to facilitate global adoption, encouraging collaboration among educators and developers to continuously enhance the system's functionality and expand its reach.

#### 7.4 | Cross-Disciplinary Technology Integration

To further elevate the system's capabilities, future versions can integrate advanced technologies such as virtual reality and augmented reality. For example, virtual laboratories can allow students to conduct scientific experiments with the guidance of digital teachers, while augmented reality spaces can facilitate hands-on skill development in real-world contexts. Additionally, incorporating technologies such as speech recognition and emotion analysis can improve the system's effectiveness in understanding and responding to learners' needs, offering a more personalized and adaptive learning experience.

## 8 Conclusions

We proposed a novel framework for creating an interactive pre-recorded lecture system deployed through a personalized digital teacher. By harnessing the power of fine-tuned LLMs for resource understanding and text generation, combined with advanced techniques for constructing lifelike digital teachers, the system delivers engaging and adaptive educational experiences. Beyond traditional lecture formats, the integration of real-time Q&A capabilities and empathetic interactions offers a more dynamic and supportive learning environment.

Case studies were carried out to demonstrate our system's potential to enhance engagement, improve learning outcomes, and provide support tailored to individual learners. An important consideration, however, is that while the proposed approach addresses some limitations of existing pre-recorded lecture systems, several areas for future exploration remain, including longitudinal learning analytics, ethical considerations, expanded applications, and cross-disciplinary technological integration. As educational technology continues to evolve, this work represents a step forward in bridging the gap between static online learning resources and fully interactive, adaptive educational experiences. By building on the strengths of AI and digital human technology, we envision a future where personalized education is accessible to learners across diverse contexts, empowering individuals and democratizing knowledge on a global scale.

**Acknowledgments** This research was partially supported by the National Natural Science Foundation of China (Grant Nos. 62337001 and 62406303), the Key Technologies R&D Program of Anhui Province (Grant No. 202423k09020039), and the Fundamental Research Funds for the Central Universities.

**Authors Contributions** Qi Liu conceptualized and designed the research, developed the overall methodology, led the project, analyzed the results, and wrote the manuscript. Yunhao Sha contributed to the design and development of the digital human framework and assisted with the implementation and testing of interactive features. Kai Zhang was responsible for data collection and analysis, playing a key role in optimizing the system for personalized interactions. Zhenya Huang contributed to the development of machine learning models for personalized learning experiences and helped with the system's evaluation. Linbo Zhu led the integration of the digital human system with recorded courses, designing and implementing the user interface. Junyu Lu contributed to the creation of personalized content features and user interaction protocols and assisted with user testing. Yu Su provided critical insights into the learning experience, particularly regarding user engagement and the pedagogical effectiveness of virtual human interaction, and contributed to manuscript editing. All authors reviewed and approved the final published version of the study, ensuring the accuracy and scientific integrity of the research content.

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Ethics Statements** The authors declare that their Institutional Ethics Committee confirmed that no ethical review was required for this study. Written informed consent for participation was not required because all participants' data was anonymized before the statistical analyses were done.

**Data Availability Statements** The authors confirm that all data generated or analysed during this study are included in this published article.

## References

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altschmidt, J., Altman, S., Anadkat, S., & et al. (2023). GPT-4 technical report. *arXiv Preprint*, arXiv:2303.08774.
- Burden, D., & Savin-Baden, M. (2019). *Virtual humans: Today and tomorrow*. New York: Chapman and Hall/CRC.
- Chopin, B., Tang, H., Otberdout, N., Daoudi, M., & Sebe, N. (2023). Interaction transformer for human reaction generation. *IEEE Transactions on Multimedia*, 25, 8842–8854.
- Eapen, J., & Adhithyan, V. S. (2023). Personalization and customization of LLM responses. *International Journal of Research Publication and Reviews*, 4(12), 2617–2627.

- Fiedler, M. L., Wolf, E., Döllinger, N., Botsch, M., Latoschik, M. E., & Wienrich, C. (2023). Embodiment and personalization for self-identification with virtual humans. In: *Proceedings of 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*. Piscataway: IEEE, 799–800.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694.
- Hadgu, R., Huynh, S., & Gopalan, C. (2016). The use of pre-recorded lectures on student performance in physiology. *Journal of Curriculum and Teaching*, 5(1), 105–112.
- Hu, B., Zheng, L., Zhu, J., Ding, L., Wang, Y., & Gu, X. (2024). Teaching plan generation and evaluation with GPT-4: Unleashing the potential of LLM in instructional design. *IEEE Transactions on Learning Technologies*, 17, 1445–1459.
- Ieronutti, L., & Chittaro, L. (2007). Employing virtual humans for education and training in X3D/VRML worlds. *Computers & Education*, 49(1), 93–109.
- Izumi, K., Tanaka, H., Shidara, K., Adachi, H., Kanayama, D., Kudo, T., & Nakamura, S. (2024). Response generation for cognitive behavioral therapy with large language models: Comparative study with Socratic questioning. *arXiv Preprint*, arXiv:2401.15966.
- Jiang, F., Dong, L., Peng, Y., Wang, K., Yang, K., Pan, C., & You, X. (2025). Large AI model empowered multimodal semantic communications. *IEEE Communications Magazine*, 63(1), 76–82.
- Jiao, J. F., Afroogh, S., Xu, Y. M., & Phillips, C. (2024). Navigating LLM ethics: Advancements, challenges, and future directions. *arXiv Preprint*, arXiv:2406.18841.
- Karpouzis, K., Pantazatos, D., Taouki, J., & Meli, K. (2024). Tailoring education with GenAI: A new horizon in lesson planning. In: *Proceedings of 2024 IEEE Global Engineering Education Conference*. Piscataway: IEEE, 1–10.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., & et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Kim, E. J., & Kim, M. J. (2024). Students' evaluation of live online lectures versus pre-recorded video lectures during COVID-19: Focusing on students' study behaviors. *SAGE Open*, 14(4), 21582440241305325.
- Kimani, E., Parmar, D., Murali, P., & Bickmore, T. (2021). Sharing the load online: Virtual presentations with virtual co-presenter agents. In: *Proceedings of Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. New York: ACM, 473.
- Kumar, H., Musabirov, I., Reza, M., Shi, J. K., Wang, X. Y., Williams, J. J., Kuzminykh, A., & Liut, M. (2023). Impact of guidance and interaction strategies for LLM use on learner performance and perception. *arXiv Preprint*, arXiv:2310.13712.
- Le, K. (2022). Pre-recorded lectures, live online lectures, and student academic achievement. *Sustainability*, 14(5), 2910.
- Lieb, A., & Goel, T. (2024). Student interaction with NewtBot: An LLM-as-tutor Chatbot for secondary physics education. In: *Proceedings of Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. New York: ACM, 614.
- Liu, Q., Huang, Z. Y., Yin, Y., Chen, E. H., Xiong, H., Su, Y., & Hu, G. P. (2021). EKT: Exercise-aware knowledge tracing for student performance prediction. *IEEE Transactions on Knowledge and Data Engineering*, 33(1), 100–115.
- Liyanaage, U. P., & Ranaweera, N. D. (2023). Ethical considerations and potential risks in the deployment of large language models in diverse societal contexts. *Journal of Computational Social Dynamics*, 8(11), 15–25.
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). A comprehensive overview of large language models. *arXiv Preprint*, arXiv:2307.06435.
- Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems*, 7(3), 11.
- Rashid, M. M., Atilgan, N., Dobres, J., Day, S., Penkova, V., Küçük, M., Clapp, S. R., & Sawyer, B. D. (2024). Humanizing AI in education: A readability comparison of LLM and human-created educational content. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 10711813241261689.
- Razafinirina, M. A., Dimbisoa, W. G., & Mahatody, T. (2024). Pedagogical alignment of large language models (LLM) for personalized learning: A survey, trends and challenges. *Journal of Intelligent Learning Systems and Applications*, 16(4), 448–480.
- Ren, Y., Ruan, Y. J., Tan, X., Qin, T., Zhao, S., Zhao, Z., & Liu, T. (2019). FastSpeech: Fast, robust and controllable text to speech. *arXiv Preprint*, arXiv:1905.09263.
- Secker, J., Reimers, S., & Foley, G. (2022). From face-to-face to remote learning: What can we learn from student experiences of pre-recorded lectures in the pandemic? In: *Proceedings of 16th International Technology, Education and Development Conference*. Valencia: IATED, 1737–1743.
- Sheng, H. X., Zhou, X. R., Zhao, Y., & Zhao, G. Q. (2023). Comparing perceived cognitive load while learning online with AI chatbots, pre-recorded videos, and live lectures. In: *Proceedings of the 31st International Conference on Computers in Education*. Published online.
- Sousa, M. J., Cruz, R. & Martins, J. M. (2017) Digital learning methodologies and tools—A literature review. In: *Proceedings of the 9th International Conference on Education and New Learning Technologies*. Barcelona, 5185–5192.
- Sun, X. J., Zhang, K., Liu, Q., Bao, M. K., & Chen, Y. J. (2024). Harnessing domain insights: A prompt knowledge tuning method for aspect-based sentiment analysis. *Knowledge-Based Systems*, 298, 111975.
- Sun, Y. N., Sun, Z. Y., Wen, Y. H., Ye, S., Lv, T., Yu, M. J., Yi, R., Gao, L., & Liu, Y. J. (2023). Generation of virtual digital human for customer service industry. *Computers & Graphics*, 115, 359–370.
- Welz, L., & Lanquillon, C. (2024). Enhancing large language

- models through external domain knowledge. In: *Proceedings of the 5th International Conference on Artificial Intelligence in HCI*. Cham: Springer, 135–146.
- Wolf, E., Mal, D., Frohnapfel, V., Döllinger, N., Wenninger, S., Botsch, M., Latoschik, M. E., & Wienrich, C. (2022). Plausibility and perception of personalized virtual humans between virtual and augmented reality. In: *Proceedings of 2022 IEEE International Symposium on Mixed and Augmented Reality*, Piscataway: IEEE, 489–498.
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112.
- Zhang, K., Gong, W. Y., Liu, Q., Lu, J. Y., Bao, M. K., Liu, X. K., Zhu, L. B., & Chen, E. H. (2025). Intelligent interaction platform for personalized digital tutor: Advancing empathetic and adaptive learning experiences. In: *Companion Proceedings of the ACM on Web Conference 2025*. New York: ACM, 2947–2950.
- Zhang, K., Liu, Q., Qian, H., Xiang, B., Cui, Q., Zhou, J., & Chen, E. H. (2023). EATN: An efficient adaptive transfer network for aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 377–389.
- Zhao, W. X., Zhou, K., Li, J. Y., Tang, T. Y., Wang, X. L., Hou, Y. P., Min, Y. Q., Zhang, B. C., Zhang, J. J., Dong, Z. C., & et al. (2023). A survey of large language models. *arXiv Preprint*, arXiv:2303.18223.
- Zhen, R., Song, W. C., He, Q., Cao, J., Shi, L., & Luo, J. (2023). Human-computer interaction system: A survey of talking-head generation. *Electronics*, 12(1), 218.