Journal of Electronic Research and Application, 2026, Volume 10, Issue 5
http://ojs.bbwpublisher.com/index.php/JERA
Online ISSN: 2208-3510
Print ISSN: 2208-3502

# Development and Application of an AI Popular Science Digital Human System Based on Local Private Large Model Technology

**Yongqiang Wang**

Beijing Sizhuang Technology Development Co., Ltd., Beijing 100000, China



**Abstract:** With the continuous development of artificial intelligence technology, digital human technology offers extensive applications in science popularization education. This paper designs an AI popular science digital human system using local private large model technology, which integrates key technologies including speech recognition, natural language processing, speech synthesis, and digital human driving to enable intelligent interactive Q&A with users. The system adopts a locally deployed architecture, fine-tuned based on the Qwen large language model, and combines SenseVoice speech recognition, CosyVoice speech synthesis, and the LiveTalking digital human driving engine to build a complete popular science interaction process. The system has been put into practical use in scenarios such as science and technology festivals in primary and secondary schools and science and technology exhibition halls, which effectively improves the fun and interactivity of science popularization education and provides a new solution for cultivating scientific literacy among teenagers.

**Keywords:** Artificial intelligence; Large language model; Digital human; Science popularization education; Local deployment; Intelligent interaction

**Online publication:** Jun 29, 2026

## 1. Introduction

In the era of the digital economy, artificial intelligence has become an important force driving social development. The 14th Five-Year National Informatization Plan clearly requires accelerating the innovation and application of artificial intelligence technology and promoting the digital transformation of education [1]. As an important way to cultivate teenagers' scientific literacy, science popularization education urgently needs new technical means to improve teaching effects and interactive experience.

Traditional science popularization education mostly delivers knowledge to students in a one-way manner, lacking interactivity and fun, making it difficult to stimulate students' interest in learning. Following new breakthroughs in large language model technology, AI digital humans, as a new form of human-

computer interaction, have brought new development possibilities to science popularization education with their human-like appearance and intelligent dialogue capabilities [2]. The AI popular science digital human system introduced in this paper is an innovative application developed under such a technical background.

## 2. System architecture design

### 2.1. Business process

The business process of the AI popular science digital human consists of multiple steps. Users interact with the system via voice or touch; the front-end interaction layer receives user operation information. The ASR speech recognition module converts voice into text, and the touch screen input parsing module processes touch operations. The system generates a question ID and user answer text and transmits them to the knowledge base matching system, which accesses the popular science Q&A knowledge base, judges whether the user's answer is correct or not, and generates Prompt content for the large model. The popular science Q&A knowledge base provides standard answers and extended knowledge text. The large model interface calls the Tongyi Qianwen or DeepSeek model to process relevant information, and the processed content is transmitted to the TTS speech synthesis module, which converts text into voice. The digital human rendering engine completes lip-syncing and expression control, and finally the output system presents voice, digital human image, and on-screen text and pictures to users (**Figure 1**).

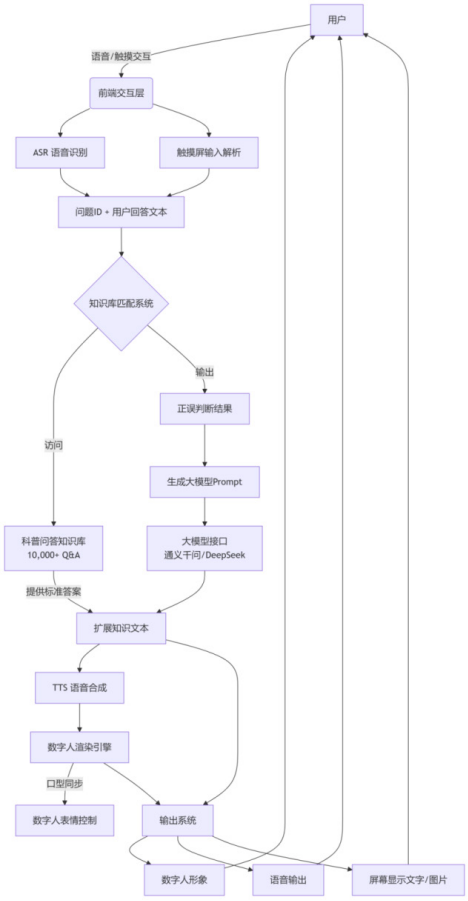


**Figure 1.** Business process of the AI popular science digital human.

## 2.2. Overall system architecture

The AI popular science digital human system adopts a layered architecture design, including a complete interaction link from front-end to back-end. It is mainly composed of three parts: front-end interaction equipment, back-end service engine, and local private large model [3]. Information is transmitted between modules through standardized interfaces to ensure system scalability and easy maintenance.

The core components of the system include: a speech recognition module that converts user voice into text; a knowledge base matching engine that performs semantic retrieval and answer matching based on a structured Q&A library; a large model gateway that calls the locally deployed Qwen model to generate extended knowledge content; a speech synthesis module that converts text into natural voice; and a digital human engine that controls virtual images to achieve lip-syncing and expression actions (**Figure 2**).

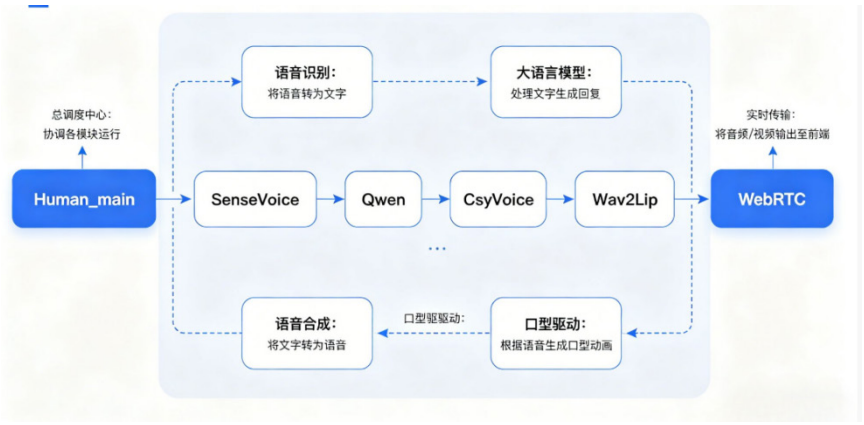


**Figure 2.** System architecture of the AI popular science digital human.

## 2.3. Technology selection

Clear selections are made for the core technologies used in the system. SenseVoice is chosen for speech recognition for its multilingual support, high accuracy, and strong environmental adaptability. Qwen with VLLM is selected as the large language model, supporting local deployment, streaming output, and low latency. CosyVoice is adopted for speech synthesis, featuring high naturalness, rich emotion, and real-time synthesis [4]. LiveTalking is used for digital human driving, realizing lip-syncing, linked expressions, and multi-state switching as shown in **Table 1**.

**Table 1.** Core technology selection of the system

| Technical Module | Selected Technology | Main Features |
|---|---|---|
| Speech Recognition | SenseVoice | Multilingual, high accuracy, strong environmental adaptability |
| Large Language Model | Qwen + VLLM | Local deployment, streaming output, low latency |
| Speech Synthesis | CosyVoice | High naturalness, rich emotion, real-time synthesis |
| Digital Human Driving | LiveTalking | Lip-syncing, linked expressions, multi-state switching |

# 3. Core key technologies

## 3.1. Speech recognition technology

The system uses SenseVoice as the speech recognition engine, which boasts multiple technical advantages:

it supports multilingual recognition with high accuracy and excellent noise suppression, adapting to complex environments such as exhibition halls [5]. It features fast response and real-time speech-to-text transcription. The model has been domain-adapted for science popularization scenarios to improve recognition accuracy for scientific terms.

### 3.2. Large language model

The system builds a knowledge generation engine based on the Qwen large language model and adopts the VLLM inference framework for efficient model deployment [6,7]. A dedicated Prompt template is designed for science popularization education scenarios, guiding the model to generate content suitable for users of corresponding age groups. The output length is controlled at about 150 words, with vivid and interesting expressions avoiding complex professional terms. Content starts with guiding phrases such as "Actually…" and "You know what…" to make the digital human's expression more approachable.

### 3.3. Digital human driving technology

The LiveTalking engine is used for digital human driving, supporting voice-driven lip-syncing technology that generates real-time lip animation corresponding to voice content [8,9]. The system also implements linked control of expressions and gestures, making the digital human image more vivid and natural [10]. The digital human can switch among multiple states including standby, interaction, and explanation modes (**Figure 3**).


**Figure 3.** Image design of the AI popular science digital human.

## 4. Application scenarios

### 4.1. Primary and secondary school science and technology festivals

In activities such as science and technology festivals in primary and secondary schools, the AI popular science digital human acts as an interactive commentator, taking the initiative to conduct popular science Q&A interactions with visiting students [11]. The system is pre-equipped with a rich science popularization question bank covering physics, chemistry, biology, astronomy and other disciplines. After students answer questions by voice, the digital human provides immediate feedback on correctness and generates extended knowledge explanations, effectively stimulating students' interest in exploring scientific knowledge (**Figure 4**).

**Figure 4.** Application scenario in campus science and technology festival.

## 4.2. Science and technology exhibition halls

In venues such as science and technology museums and museums, the AI popular science digital human can be installed on interactive display terminals to provide 24/7 intelligent explanation services for visitors. The system can customize the content of the knowledge base according to the theme of the exhibition hall to achieve targeted popular science services (**Figure 5**). The local private deployment ensures data security during use, faster system response and lower latency [12,13].


**Figure 5.** Application scenario in science and technology museum.

### 4.3. Other popular science scenarios

In addition to the main scenarios mentioned above, the system can also be used in mobile popular science exhibitions, library popular science corners, community popular science activities, youth palace science laboratories and other scenarios. The modular design allows convenient transplantation, and rapid customization and deployment according to the needs of different scenarios (**Figure 6**).


**Figure 6.** Application scenario of AI popular science into business districts (Beijing Association for Science and Technology).

## 5. System function design

### 5.1. Interactive Q&A process

The core function of the system is interactive Q&A. The complete interactive Q&A process includes multiple steps: the user wakes up the digital human via voice or button; the digital human takes the initiative to ask a popular science-related question; the user answers the question by voice; the system performs semantic matching to judge whether the user's answer is correct; the digital human provides immediate feedback and generates extended knowledge explanations. The entire process is user experience-oriented, with reasonable timeout control and exception handling mechanisms set for each step.

### 5.2. Knowledge base management

The system adopts a structured knowledge base design. Each Q&A entry includes question text, standard answer, keyword list, topic classification, difficulty level and other content. The knowledge base supports batch import and online editing, and administrators can easily maintain the content through the back-end management system. The matching engine supports two working modes: keyword matching and semantic similarity calculation, ensuring more accurate answer judgment results.

### 5.3. Interactive experience optimization

To improve the interactive experience, the system has been optimized in many aspects. An encouraging feedback strategy is adopted: when the user answers incorrectly, instead of directly saying "You're wrong", it uses phrases like "That's an interesting idea, but the answer is…". The interaction rhythm is reasonably controlled, with sufficient response time reserved for questioning, waiting, feedback and expansion. If no interaction is received within 30 seconds, the system automatically enters a sleep state to save operating resources.

## 6. Conclusion

This paper introduces an AI popular science digital human system based on local private large model technology, which integrates multiple AI technologies including speech recognition, natural language processing, speech synthesis and digital human driving to build a complete intelligent popular science interaction solution. The system has been put into practical use in scenarios such as primary and secondary school science and technology festivals and science and technology exhibition halls, verifying its technical feasibility and practical application value. In future work, the R&D team will continue to optimize the system in several aspects: first, enrich the digital human image library to support more personalized customization options; second, expand the coverage of the knowledge base with more disciplinary content; third, add multimodal interaction capabilities to support display forms such as images and videos; fourth, explore integration with VR/AR technology to create a more immersive popular science experience. We believe that AI digital human technology will play an increasingly important role in the field of science popularization education and contribute more to cultivating teenagers' scientific literacy and innovative spirit.

## Disclosure statement

The author declares no conflict of interest.

## References


[1] National Informatization Planning Working Group, 2021, 14th Five-Year National Informatization Plan.

[2] Brown T, Mann B, Ryder N, et al., 2020, Language Models are Few-Shot Learners, Advances in Neural Information Processing Systems, 1877–1901.

[3] Alibaba Cloud, 2023, Qwen Large Language Model Technical Report.

[4] Alibaba Group, 2024, CosyVoice Speech Synthesis Technical Documentation.

[5] Alibaba Group, 2024, SenseVoice Speech Recognition White Paper.

[6] OpenAI, 2023, GPT-4 technical report, arXiv preprint arXiv:2303.08774.

[7] Kwon W, Li Z, Zhuang S, et al., 2023, Efficient Memory Management for Large Language Model Serving with Pagedattention, SOSP, 611–626.

[8] LiveTalking Open Source Community, 2024, LiveTalking Digital Human Driving Engine Documentation.

[9] Prajwal K, Mukhopadhyay R, Namboodiri V, et al., 2020, A Lip Sync Expert is all you Need for Speech to Lip Generation in the Wild, ACM MM, 484–492.

[10] Yang F, Zhang W, 2024, Research on Deep Learning-Based Voice-Driven Digital Human Technology. Journal of Computer Research and Development, 61(2): 312–325.

[11] China Association for Science and Technology, 2021, National Scientific Literacy Action Plan (2021–2035).

[12] Ministry of Education, 2018, Education Informatization 2.0 Action Plan.

[13] Liu Y, Zhao W, 2023, Analysis on Development Status and Trends of Digital Human Technology. Journal of Computer-Aided Design & Computer Graphics, 35(8): 1156–1168.