applied system innovation

MDPI

*Article*

# Integration of AI Content Generation-Enabled Virtual Museums into University History Education

**Shirong Tan, Yuchun Liu * and Lei Wang**

School of Marxism, Wuhan University of Science and Technology, Wuhan 430000, China; 15573481091@163.com (S.T.); candywang@wust.edu.cn (L.W.)
* Correspondence: liuyuchun@wust.edu.cn

**Abstract**

Traditional approaches to university-level history education often fail to provide immersive and interactive environments that foster deep cognitive engagement. To address these limitations, we developed an AI-enabled virtual museum system that integrates AI-generated content with knowledge graphs through a multi-layered architecture. The system architecture follows a three-tier framework: a front-end interaction layer (Unity/Unreal Engine) for real-time user engagement, a core service layer for intelligent event scheduling and response control (Chat General Language Model/Stable Diffusion), and a data and model layer (My Structured Query Language/MongoDB) to provide structured knowledge. To evaluate the system's effectiveness, a four-week controlled experiment was conducted with 83 university students. The experimental group using the AI virtual museum showed a significantly higher mean post-test score (84.5 $\pm$ 6.8) than that of the control group (71.6 $\pm$ 7.9), with statistical significance at $p < 0.001$, starting from nearly identical baseline scores (61.2 and 60.4 for the experimental and control groups). Correlation analysis was conducted to identify scenario simulations (r = 0.59) and deep inquiry tasks (r = 0.54) as key drivers of learning mastery. By aligning advanced system engineering with educational theory, the results of this study offer a solution for high-fidelity, intelligent digital educational platforms, proposing a validated model for integrated system innovation in education.

**Keywords:** AICG technology; modern Chinese history; knowledge graph; virtual museum

Check for updates

Academic Editor: Claudio Zunino

Received: 7 February 2026
Revised: 10 March 2026
Accepted: 13 March 2026
Published: 18 March 2026

**Copyright:** © 2026 by the authors. Published by MDPI on behalf of the International Institute of Knowledge Innovation and Invention. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

## 1. Introduction

The rapid advancement of AI content generation (AICG) technology is reshaping virtual presentations, digital culture, and educational practices. AICG automatically generates text, images, audio, and 3D content, demonstrating both efficiency and creative potential in semantic understanding, image reconstruction, and scene generation [1]. These capabilities align with the instructional demands of the compulsory course 'Outline of Modern Chinese History in Chinese' in higher education, addressing the need for tangible historical evidence and perceptible historical contexts. As a result, AICG offers an effective solution to long-standing challenges in traditional history teaching, such as limited access to historical artifacts and the difficulty of restoring historical scenes.

Previous research on virtual display technologies and intelligent instructional assistance in history education has primarily focused on static content presentation and one-way interaction. Current systems lack mechanisms for real-time feedback and adaptive adjustment to learner behavior [2]. To address these limitations, we developed an intelligent

interactive virtual museum through the coordination of AICG and knowledge graphs. By continuously collecting learner interaction data within the virtual museum, the system generates AICG content and adapts knowledge graph pathways, forming a closed-loop cycle of perception, generation, and feedback. This real-time linkage mechanism constitutes the core technical contribution of the study and is systematically validated through system implementation and experimental analysis. To enhance instructional effectiveness, we integrated a virtual museum into 'Outline of Modern Chinese History', which is a compulsory subject at many Chinese universities. The subject is characterized by extensive content coverage and large student enrollment, which places increasing pressure on educational innovation. The developed system introduces AICG-generated multimodal content that presents historical artifacts, events, and documents in an integrated and spatialized manner into the course. This approach enhances visual clarity and immersion, thereby increasing student engagement and promoting deeper understanding of the course material. The novelty of this study lies in its integration of AICG with knowledge graphs to establish a real-time closed-loop interaction mechanism. Unlike previous systems that relied on static displays or one-way interaction, our approach dynamically adapts to learner behavior, thereby enhancing instructional effectiveness. This contribution addresses a critical gap in intelligent educational technologies, as highlighted in recent reviews of AI-assisted history education [3].

The AI-enabled museum was designed based on the constructivist learning theory, which posits that learners actively construct knowledge rather than passively absorbing information. Through interactions with AI agents representing historical personas, students engage in experiential learning, moving from abstract historical facts to concrete, immersive experiences [4]. This pedagogical framework ensures that the immersion within the virtual reality (VR) system meets a cognitive purpose, enabling scaffolded inquiry where AI provides real-time feedback for the student's zone of proximal development.

## 2. AI Virtual Museum

We developed an AI virtual museum that enables real-time, interaction-driven intelligent content generation. By integrating AICG-based content generation with the structured semantic organization of a knowledge graph, the system dynamically generates and presents cultural relic information. The system integrates a user interaction perception mechanism to collect learners' browsing, operational, and task-related behaviors in real time. The interaction data collected are used for content generation and knowledge recommendation, forming a closed-loop intelligent interaction framework.

### *2.1. Hardware Deployment*

#### 2.1.1. Interactive Terminal Devices

To accommodate diverse teaching scenarios, the system comprises multiple types of interactive terminals. VR head-mounted displays (VR-HMDs) support immersive 3D artifact exploration and historical scene observation. Featuring 4K resolution, a 90 Hz refresh rate, and high processing power, these devices ensure stable rendering in complex 3D environments [5]. To minimize cybersickness, the system was optimized to maintain a consistent frame rate of 90 frames per second. We implemented a decoupled rendering architecture where AI-generated content is processed on an external edge server (NVIDIA RTX 3090, sourced from Wuhan, China). During AI response generation, the user's field of view remains static or shows a neutral interface, preventing vestibular-ocular conflicts that can cause nausea during high-latency events [6].

Touchscreen all-in-one terminals allow synchronous multi-user browsing and interaction, supporting operations such as zooming, rotating, and annotating 3D models. Desktop

and mobile devices are used to access the system through web-based or lightweight client applications, allowing flexible use regardless of time and location. Beyond content presentation, these terminals collect real-time interaction data, including click actions, viewpoint changes, dwell time, navigation paths, and task submissions, which serve as real-time input data for the event-driven mechanism and subsequent AICG content generation and knowledge-graph linkage adjustments.

#### 2.1.2. Data Storage and Computing

The virtual museum system involves 3D models, high-resolution images, historical documents, and large-scale user behavior logs, requiring considerable storage and computing resources. The system adopts a high-throughput storage architecture that includes solid-state drives and non-volatile memory express technologies to reduce latency in 3D asset loading and model access. Data security and system recoverability are ensured through a redundant array of independent/inexpensive disks and cloud-based backup solutions [7].

For computation, the system is equipped with graphics processing units (GPUs) for the real-time operation of AICG models. For fast computing, a hybrid architecture combining GPU servers and cloud-based computing power is employed. With distributed scheduling, the system processes 3D rendering, semantic content generation, speech synthesis, and knowledge association in parallel [8]. This architecture ensures low response latency and stable system performance even under high-frequency event triggering and concurrent multi-user access, thereby providing robust engineering support for the real-time interaction and linkage mechanisms.

#### 2.1.3. Multi-Device Compatibility

To ensure consistent operation of the event-driven mechanism across different terminals, a cross-platform development strategy and a unified interface architecture are adopted in the system. Unity, Unreal Engine, and Web Graphics Library (WebGL)/WebGPU are used for unified access and functional consistency across multiple VR devices, touch-screen terminals, desktop systems, and mobile devices [9]. Unity 2022.3 LTS is used as the primary development engine owing to its stable support of Open Extended Reality and its ability to handle asynchronous Representational State Transfer application programming interface (API) calls to Large Language Models (LLMs) and image generation servers without interrupting the main rendering thread [6].

The system is compatible with Windows, Macintosh Operating System, Android, iPhone Operating System, and mainstream VR operating systems, ensuring stable performance in diverse hardware environments. AI functional modules are available through standardized APIs. Regardless of terminal type, the system performs the same AICG content generation, 3D rendering, and knowledge graph linkage services [10]. Such multi-terminal, consistent-access deployment ensures the continuity and scalability of the event-driven mechanism in complex and varied applications.

#### 2.1.4. Scalability and Maintenance Strategy

To ensure sustainability, the system adopted a microservices architecture to update individual content modules without system-wide downtime. For resource management in high-concurrency scenarios, a load balancer distributes requests across the hybrid GPU-cloud cluster. This architecture supports automated content pipelines, enabling instructors to upload new 3D assets and update the knowledge graph with minimal technical intervention [11].

### 2.2. Software Architecture

The system adopts a three-layer software architecture: a front-end interaction layer, core service layer, and data and model layer (Figure 1). The system architecture is organized into three hierarchical layers, each designed to support an immersive and responsive virtual museum experience.

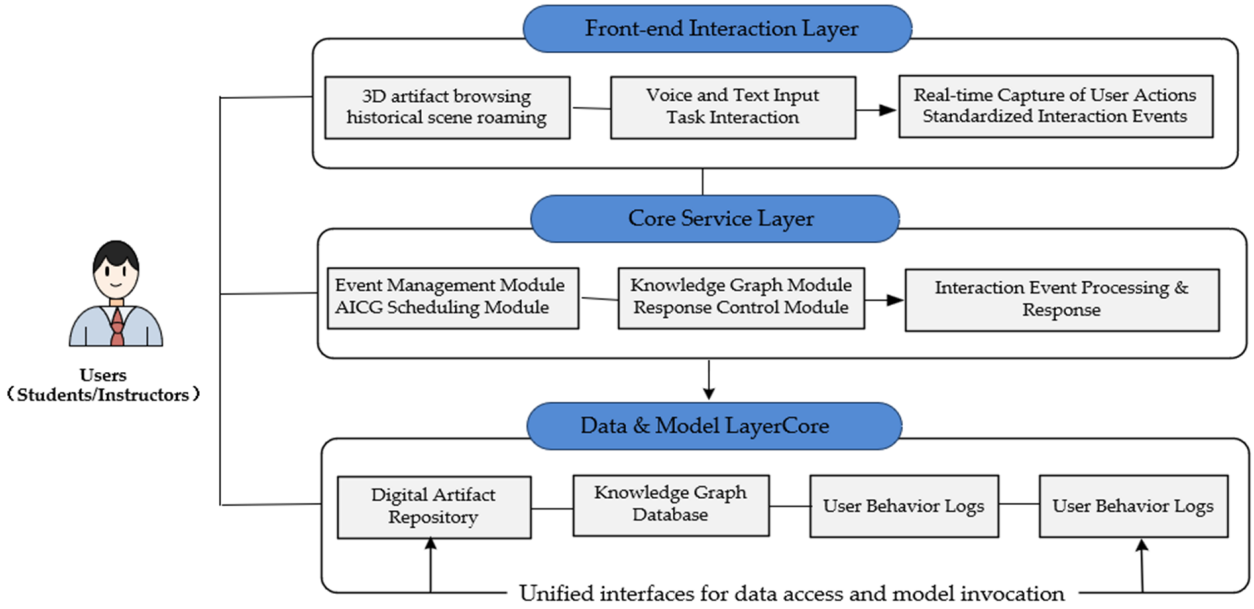



**Figure 1.** System architecture.

#### 2.2.1. Front-End Interaction Layer

This layer serves as the primary interface for students and instructors. It handles the visual rendering and real-time engagement through 3D artifact browsing and historical scene roaming. In addition to visual display, this layer processes multimodal voice and text inputs for task-oriented interactions. A main function of this layer is the real-time capture of user actions, which utilizes a unified event-listening mechanism to track clicks, navigation paths, and viewpoint changes. These actions are then transformed into Standardized Interaction Events and transmitted to the backend for processing.

#### 2.2.2. Core Service Layer

Functioning as the operational core, this layer parses and orchestrates system logic. It is composed of the following specialized modules.

- The event management and the AICG scheduling modules prioritize incoming events and trigger the AICC pipelines;
- The knowledge graph and the response modules link user queries to historical data and determine the most appropriate feedback, whether visual, textual, or behavioral;
- The interaction event processing and response module ensures that the workflow, from content generation to knowledge association, is executed seamlessly before a final response is sent back to the user.

#### 2.2.3. Data and Model Layer

This foundational layer provides infrastructure for data persistence and model execution. It maintains the digital artifact repository (3D assets) and the knowledge graph database (relational historical data). The user behavior log captures raw interaction data to refine future system responses and track learning progress [12]. By providing standardized

APIs for data access and model invocation, this layer allows the core service layer to query complex datasets and call AICG models efficiently, ensuring stability during high-frequency concurrent access [13].

The system employs a specialized technology stack to integrate immersive front-end experiences with intelligent back-end processing and ensure seamless interaction between virtual scene rendering, artificial intelligence algorithms, and secure data management (Table 1).

**Table 1.** Front-end development tools of the system.

| Purpose | Adopted Technology | Application |
| --- | --- | --- |
| Front-end development | Unity and Unreal Engine | Virtual scene modeling, 3D interaction, and viewpoint switching |
| Back-end development | Spring Boot and MyBatis | API development, request handling, database interaction, and access control |
| AI algorithm libraries | TensorFlow | Artifact image feature extraction and key-point detection for 3D models |
| | PyTorch3.10 | Scene generation optimization and model inference acceleration |
| | ChatGLM | Historical question answering, dialogue generation, and AI-guided virtual tours |
| | Stable Diffusion | Dynamic historical scene generation and event atmosphere reconstruction |
| Data storage | MySQL + MongoDB | Storage of educational data and digitized cultural artifact resources |
| Security | AES + RBAC | Storage of educational data and digitized cultural artifact resources |

For front-end development, Unity and Unreal Engine are used to create high-fidelity virtual scene models, enable interactive engagement with three-dimensional artifacts, and support smooth viewpoint switching. On the back end, Spring Boot and MyBatis are employed for service interfaces, request handling, and access control. These frameworks collectively establish a robust API layer that enables reliable database interaction and system scalability.

The intelligence of the virtual museum is supported by a multi-modal suite of AI models. For computer vision, TensorFlow is used for feature extraction and key-point detection of 3D models, while PyTorch optimizes scene generation and accelerates model inference. For generative tasks, the ChatGLM model is integrated to provide historical question answering and AI-guided virtual tours. Complementing this, Stable Diffusion is used to generate dynamic historical scenes and reconstruct authentic event atmospheres, thereby enhancing the immersive quality of the museum environment.

To accommodate diverse data types, the system employs a hybrid storage strategy: MySQL for structured educational data and MongoDB for digitized cultural artifact resources. This combination ensures efficient handling of heterogeneous datasets and supports both transactional and document-oriented queries. Data integrity and privacy are ensured using the Advanced Encryption Standard (AES) for data at rest. In addition, role-based access control (RBAC) is applied to regulate user permissions, ensuring that access to sensitive resources is restricted according to predefined user roles. Beyond technical encryption, the platform was designed to comply with the Personal Information Protection Law. All user behavior logs were anonymized at the source using a unique identifier hash, ensuring that individual student identities are decoupled from their learning analytics [14].

To ensure the generated content remains academically rigorous, the system utilizes few-shot prompting and negative constraints [15]. To prevent historical anachronisms, the

system employs temporal entity linking. Each artifact and event in the knowledge graph is tagged with a valid time attribute. The AI engine cross-references the student's current exploration against knowledge graph metadata. If a student's query involves entities outside that timeframe, the system triggers a temporal guardrail, explaining that the event or item does not yet exist in the current historical context [16].

### 2.3. Functional Modules

The core service layer functions as the pedagogical engine, shifting from static scripts to dynamic, AI-driven interactions. The system integrates LLMs to facilitate scaffolded inquiry, processing natural language inputs through a prompt-engineering framework to ensure historical accuracy [15]. This design enables AI to function as a knowledgeable guide within the student's zone of proximal development, offering context that helps students reach analytical conclusions rather than simply receiving raw data [6]. The intelligent interaction design is supported by multimodal voice- and text-based interfaces, enabling historical event tracing, artifact analysis, and scenario-based decision simulation (Figure 2).

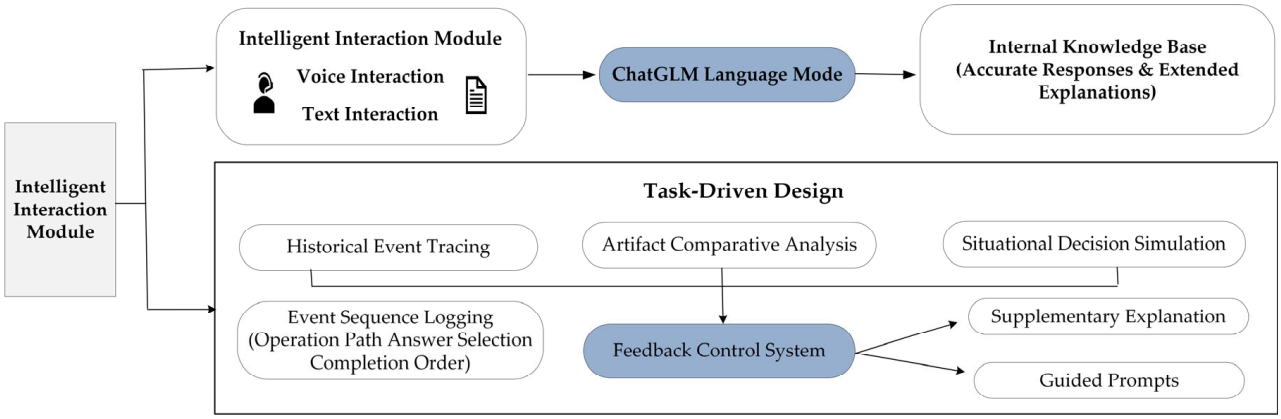



**Figure 2.** Intelligent interaction module.

In addition to textual descriptions, the system employs a diffusion model for real-time visual generation. When a student inquires about a specific historical context not included in 3D assets, the core layer conducts a localized generation task [17]. This integration of AICG supports experiential learning by delivering immediate, high-fidelity visual feedback to student queries, bridging the gap between theoretical historical knowledge and visual comprehension [18]. The content presentation module further enhances immersion through 3D reconstruction, viewpoint switching, and dynamic interaction with artifacts and reconstructed historical scenes (Figure 3).

Intelligent scheduling is supported by a domain knowledge graph, ensuring that all AI-generated responses are grounded in verified historical causality [16]. We employ a retrieval-augmented generation architecture. Before AI generates a response, the system queries a domain knowledge graph for relevant triples. These triples are injected into the prompt as 'ground truth'. The system logic is programmed with a Knowledge-Override protocol. If the LLM's output entities do not match the symbolic entities in the knowledge graph, the system rejects LLM output and forces a retrieval-based response to ensure historical accuracy [19]. This architecture ensures that the virtual characters' responses remain anchored in verified historical data while maintaining natural conversational flow. By mapping relationships between events, figures, and artifacts, the system enables students to identify cause-and-effect chains. For example, if a student interacts with a specific tool, the system suggests related political or economic events of that era, fostering higher-order thinking skills such as analysis and evaluation, which are critical for university-level

history education [20]. The knowledge graph module dynamically adjusts to user interactions, extending associations and refining hierarchies to present historical information as interconnected networks rather than isolated points (Figure 4).

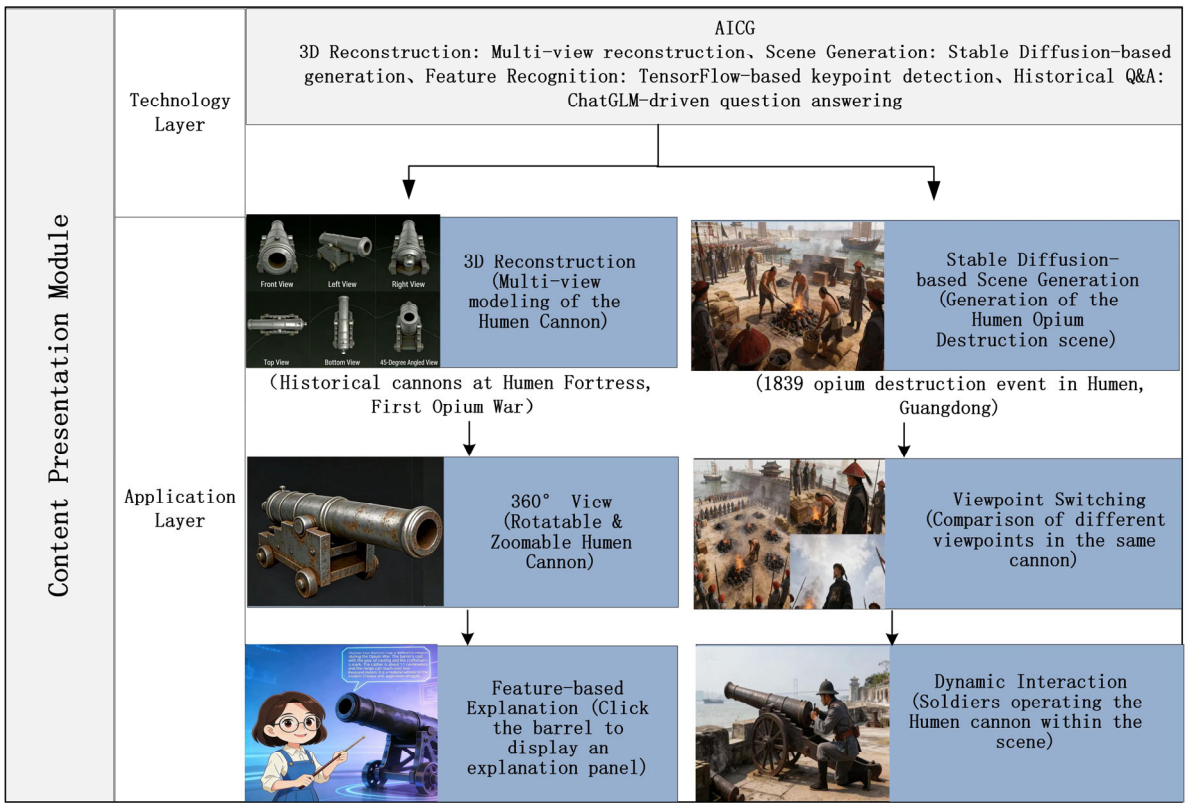



**Figure 3.** Content presentation module showing a human cannon as an example.

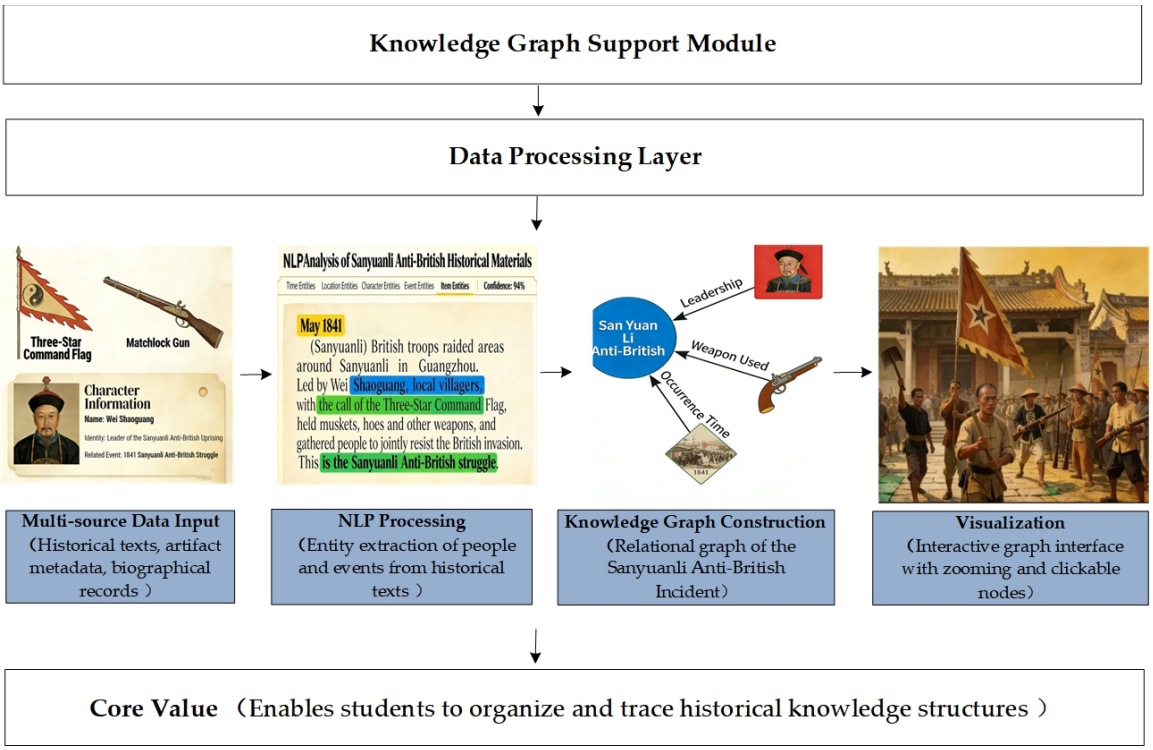



**Figure 4.** Knowledge graph module showing the Sanyuanli anti-British movement case.

Bayesian Logic for Conflict Resolution

Data conflicts between the LLM and the knowledge graph are resolved by using a Bayesian Conflict Resolution framework. The framework enables the system to calculate the probability of accuracy based on source reliability. The knowledge graph is assigned a high prior probability ($P(fact)$ = 0.95), while unverified AI generation is assigned $P(fact)$ = 0.40. The posterior probability $P(fact|source)$ is calculated as follows.

$$P(fact|source) = \frac{P(source|fact)\ P(fact)}{P(source)} \quad (1)$$

Here, $P(fact|source)$ is the probability that a specific historical fact is true, given that a particular source (the Knowledge Graph or the AI) provided it; $P(fact)$ is the probability that a specific historical fact is true; $P(source)$ is the total probability that the source reports the fact under all possible circumstances; and $P(source|fact)$ is the probability that the source reports this fact, assuming the fact is actually true. $P(fact|source)$ is the confidence score. If this value is high (>0.85), the system displays the information to the student. If it is low, the system rejects the AI's answer and triggers a self-correction loop. $P(fact)$ represents the trust level of the database. Since the knowledge graph is curated by experts, its $P(fact)$ is set high (0.95). Since the LLM can hallucinate, its $P(fact)$ is set lower (0.40) to ensure the system trusts the knowledge graph more than the AI. $P(source)$ is used as a normalization constant to ensure the final probability stays between 0 and 1. It accounts for how often the source makes claims in general. $P(source|fact)$ is used as a measure of the source's consistency. If the knowledge graph matches academic textbooks 99% of the time, the likelihood is 0.99. If the AI often mixes up dates, its likelihood becomes lower [18].

## 3. Sample Collection

The AI virtual museum system was deployed for four weeks from 1 to 28 September 2025. In the system, digital artifacts and instructional scenarios served as data collection points to trigger interaction and were used as objective sources for subsequent experimental analysis based on the integration of AICG and knowledge graphs [17].

### *3.1. Design of Teaching Units*

In the virtual museum system, digital artifacts serve as the central nodes for interaction and data representation. Artifacts were selected for historical significance, curricular relevance, and feasibility of digital processing. These considerations ensure that the chosen artifacts encompass key stages of modern Chinese history and meet engineering requirements, including 3D reconstruction, semantic recognition, and the triggering of interactive events within the system [16]. Based on these criteria, eight representative artifacts were selected to constitute the core sample units of the virtual museum: the Humen Cannon, the San Yuan Li anti-British command flag, the petition list of the Tibetan local government opposing British aggression, and the Five-Colored Republican Flag.

#### 3.1.1. Digital Processing

In the digitization stage, the system builds upon existing museum digital resources and performs secondary processing and instructional scene reconstruction for selected artifacts. For physical artifacts such as the Humen Cannon and Ma Benzhai's command saber, the system uses 3D reconstruction models, texture enhancement, and lighting correction algorithms to improve the visibility of details. This enables users to observe traces of historical use through rotation, scaling, and close inspection. Inscription-based and documentary artifacts rely on high-resolution image resources combined with optical character recognition and image-text matching models to enable automatic text recognition

and structured storage of content. Each artifact is represented as an interactive object node, with its 3D model, image textures, textual information, and historical labels uniformly encapsulated. When users interact with an artifact, the system captures the corresponding interaction in real time and records it for subsequent experimental analysis [21].

#### 3.1.2. Selecting Cultural Artifacts

Cultural artifacts were selected based on their historical representativeness, curricular relevance, and feasibility of digitization.

First, the selected artifacts cover three major historical stages: awakening, transformation, and the emergence of a new era, which correspond to the knowledge structure of modern Chinese history. Second, the artifacts possess cultural symbolism and instructional significance, enabling the intuitive presentation of the Chinese national community consciousness across different historical periods. Such artifacts include national flags, oath inscriptions, and documents related to united front practices, all of which carry symbolic meaning. Third, artifacts are supported by historical sources and are embedded within instructional contexts. Each selected artifact is linked to historical events, key figures, and theoretical concepts addressed in the curriculum. This linkage offers a coherent artifact–event–knowledge point in the virtual museum system.

### *3.2. Teaching Scenario and System Interaction*

#### 3.2.1. Knowledge Construction and Inquiry-Based Learning Scenarios

The system constructs a knowledge graph related to the artifacts, integrating fragmented information distributed across different spaces and curriculum units. By analyzing relationships among artifacts, historical events, and contextual meanings, the system constructs a structured knowledge network based on the history of foreign imperialist aggression, the history of the Chinese people's resistance, and the history of China's modernization efforts. When students learn about the formation of the Chinese nation as a self-conscious entity, the knowledge graph links relevant artifacts, documentary records, historical photographs, and archival materials, enabling the multidimensional presentation of related cultural relics and historical sources. Each artifact presents its physical attributes and historical background, including time, location, key figures, and ideological context, which are connected to related events. By browsing 3D artifact models and using AI-generated textual and audio explanations, students learn about relationships among historical events and figures along the knowledge graph. The system visualizes individual learning paths, enabling students to track their learning trajectories. This process supports the integration of perceptual experience, logical understanding, and conceptual construction of the evolution of modern Chinese history [22].

#### 3.2.2. Practice-Oriented Application and Feedback Scenarios

Using the system, students engage in immersive historical scenarios through role-based participation. In virtual scenes such as the Sanyuanli anti-British movement case, the Humen Opium Destruction, and the Five-Race Republic Flag, students reconstruct historical events and engage in simulations that involve interaction with digital artifacts and documents. As a result, users can observe historical use and documentary evidence. Based on students' interactions, task completion status, and behavioral data, the system generates personalized learning reports, including task accuracy rates, mastery of key historical concepts, and unexplored knowledge nodes. Within simulation-based activities, students actively analyze historical evidence, reconstruct event sequences, and make decisions about how past events unfolded and why they mattered. They evaluate competing interpretations, weigh the credibility of sources, and form reasoned judgments about historical causality and significance. AI-guided support enhances this learning process by

offering timely feedback to clarify historical logic, guiding students to relevant artifacts and documents, and suggesting pathways for extended learning. Visualization of learning paths and knowledge graph mappings allows students to reflect on their learning processes, thereby enhancing self-directed inquiry and overall effectiveness. At the same time, such data-driven insights help instructors adjust teaching strategies and optimize classroom guidance [23].

## 4. Effectiveness of AI Virtual Museum

We evaluated the effectiveness of the AI virtual museum system in the following dimensions: system operation data, interaction event indicators, and overall outcome. The results highlight the technical mechanisms that influence information acquisition, interaction, and historical knowledge construction. The evaluation criteria were standardized based on the Cognitive Domain of Bloom's Taxonomy, targeting analysis (identifying causal links) and evaluation (judging source credibility). Historical literacy was assessed in dimensions of chronological thinking, source analysis, historical comprehension, and causal reasoning [4].

### *4.1. Data Collection and Analysis*

We surveyed students in two classes of comparable size and similar academic backgrounds during the same teaching period. The students learned the thematic unit "The Formation of the Self-Aware Chinese Nation" from the course Outline of Modern Chinese History. A total of 83 students were recruited from undergraduate history courses. Using a randomized controlled trial design, the participants were assigned to the experimental group ($n$ = 41; 51% were female, and the mean age was 20.2 years) and the control group ($n$ = 42; 49% were female, and the mean age was 20.5 years). To ensure consistency, the same instructor delivered the pre-lecture and post-session debriefing to both groups. This study was approved by the Institutional Review Board (IRB) of Wuhan University of Science and Technology, and written informed consent was obtained from all participants before data collection.

The AI virtual museum system was deployed for four weeks. In this period, the system automatically recorded and stored all user interaction data in the backend for subsequent analysis. Three types of data were collected as follows.

- Learning outcomes as external validation indicators of instructional effectiveness;
- Interaction events automatically captured by the system during student engagement with artifacts;
- Qualitative feedback obtained through semi-structured interviews with a subset of the participants.

Interaction event data were treated as independent variables, while learning outcomes served as dependent variables. To evaluate system effectiveness, independent-samples $t$-tests were conducted to compare learning outcomes between the intervention and non-intervention groups. Statistical significance was used to determine the impact of system integration on student performance. Pearson correlation coefficients were also calculated to examine relationships between different types of interaction events and learning performance, thereby assessing the influence of event-driven mechanisms on system effectiveness.

Qualitative data were gathered in semi-structured interviews (15–20 min) with 15 purposively selected students from the experimental group. Semi-structured interviews were conducted to examine how the system's interaction mechanisms affected user behavior. The interview protocol focused on perceived agency and historical empathy. Transcripts were analyzed using thematic analysis. Two independent researchers coded the data. The coded data yielded an inter-coder reliability (Cohen's Kappa) of 0.84, indicating

strong consistency in identifying themes such as emotional connection to artifacts and AI narrative reliability [24].

### 4.2. System Effectiveness

The learning outcomes of the two groups were compared before and after system employment. For the assessment of learning outcome differences, pre- and post-tests were conducted for the two groups. The test results were reported using a 100-point scale and categorized into high (≥90), upper-middle (80–89), middle (60–70), and low (<60) score levels. To ensure the reliability of the survey instrument used to assess system acceptance and perceived learning gains, we calculated Cronbach's $\alpha$. The overall scale demonstrated high internal consistency ($\alpha = 0.88$), indicating that the items reliably measured the intended constructs. The post-test results showed that the score of the experimental group ($84.5 \pm 6.8$) was significantly higher than that of the control group ($71.6 \pm 7.9$) ($p < 0.001$, $t = 6.42$). The calculated effect size (Cohen's d) was 1.75, with a 95% confidence interval of [1.25, 2.25]. This represents a large effect size according to Cohen's conventions, indicating that the AI-integrated virtual museum substantially improved student performance compared with traditional instructional methods.

Before the system employment, there was no statistically significant difference between the two groups ($t = 0.394$, $p = 0.691 > 0.05$) (Table 2). This suggests that the two groups were identical in terms of baseline cognitive level, providing a valid prerequisite for subsequent analysis of system effectiveness.

**Table 2.** Comparison of pre-test results.

| Group | Number of Students | Mean Score | Standard Deviation | *t* | *p* |
|---|---|---|---|---|---|
| Experimental group | 42 | 61.2 | 8.4 | 0.394 | 0.691 |
| Control group | 41 | 60.4 | 7.9 | | |

The post-test results showed that the experimental group scored significantly higher than the control group ($p < 0.01$) (Table 3). The result indicates that the AI virtual museum system had a significant positive effect on students' learning outcomes.

**Table 3.** Comparison of Samples After System Intervention.

| Group | Number of Students | Mean Score | Standard Deviation | *t* | *p* |
|---|---|---|---|---|---|
| Experimental group | 42 | 84.5 | 6.8 | 6.42 | 0.001 |
| Control group | 41 | 71.6 | 7.9 | | |

The distribution of scores underscores the system's effect on learning outcomes (Table 4). In the post-test, 73.1% of students in the system-intervention group scored in the high or upper-middle levels, compared with only 41.6% in the non-intervention group. The control group remained primarily at the middle level (54.3%), whereas the experimental group demonstrated improved scores. These results indicate that the AI virtual museum system is effective in fostering integrative analytical abilities and higher-level understanding, rather than merely enhancing basic knowledge acquisition.

**Table 4.** Learning outcomes (scores) of the experimental and control groups.

| Score Level | Percentage in Each Score Level | |
|---|---|---|
| | Experimental Group | Control Group |
| High (≥90) | 28.0% | 10.0% |
| Upper-middle (80–89) | 45.1% | 31.6% |
| Middle (60–70) | 25.6% | 54.3% |
| Low (<60) | 1.2% | 4.1% |

Based on the interaction logs automatically recorded by the system, correlation analysis was conducted between interaction event indicators and learning outcomes of the experimental group. Artifact viewing time was measured in minutes to capture the total time a user spent actively examining 3D artifacts. It reflects the system's capacity to support deeper information acquisition. Participation in in-depth inquiry tasks was measured by the number of times students engaged with complex exploratory tasks. Scenario simulation completion was counted by the number of scenarios completed. Extended reading clicks were measured by the number of times a user accessed supplementary historical data. Multimedia interaction usage was measured by the frequency of engagement with audio, video, or interactive media elements. Discussion and feedback triggers were measured by the number of times students initiated communication or feedback within the system. The basic task completion rate was calculated as the percentage of predefined instructional objectives successfully achieved (89%).

All interaction event indicators showed positive correlations with learning outcomes. The number of completed situational simulations ($r = 0.59$, $p < 0.001$) and the frequency of deep inquiry task engagement ($r = 0.54$, $p < 0.001$) exhibited the strongest correlations with learning outcomes, suggesting that event-driven interaction enhanced learning outcomes. Artifact viewing time ($r = 0.48$) and basic task completion rate ($r = 0.50$) supported deeper information acquisition and structured learning guidance. In contrast, extended reading behaviors and multimedia interaction usage showed weaker correlations and functioned as auxiliary facilitators, although maintaining stable positive correlations (Table 5).

**Table 5.** Interaction event indicators and correlation with learning outcomes.

| Indicator | Measurement Results in Average | Correlation Coefficient with Learning Outcome (*r*) | Significance Level |
|---|---|---|---|
| Artifact viewing time | 24.3 min | 0.48 | $p < 0.01$ |
| Basic task completion rate | 89% | 0.50 | $p < 0.01$ |
| Participation in in-depth inquiry tasks | 16 times | 0.54 | $p < 0.001$ |
| Scenario Simulation Completion Count | 4 scenarios | 0.59 | $p < 0.001$ |
| Extended reading clicks | 12 times | 0.46 | $p < 0.01$ |
| Multimedia interaction usage | 7 times | 0.42 | $p < 0.01$ |
| Discussion and feedback triggers | 5 times | 0.36 | $p < 0.05$ |

To examine the factors contributing to the observed improvement in learning outcomes and to evaluate the effectiveness of the AI virtual museum, we conducted a thematic coding analysis of interview transcripts from 25 students in the experimental group. Through open coding and axial coding, three thematic categories were identified: acceptance of the virtual museum, perceived learning gains, and identified issues with suggested improvements (Table 6).

**Table 6.** Thematic coding results of student feedback on the virtual museum.

| Category | Item | Number of Respondents | Student Feedback |
|---|---|---|---|
| Acceptance of the virtual museum | Enhanced immersion through virtual scenes | 18 | Strong sense of being immersed in history |
| | Improved understanding supported by AI-guided explanations | 15 | AI-based artifact interpretation saves time and improves memorability |
| | Easy and intuitive system operation | 12 | Simple operation, easier to understand than PowerPoint-based teaching |
| Perceived learning gains | Combined scenes and artifacts supported deeper historical understanding | 17 | Scenes help connect historical causes and consequences |
| | Visualized content strengthened memory | 13 | Key meanings remembered more quickly |
| | Inquiry-based tasks increased learning motivation | 13 | Inquiry-based tasks stimulate exploration motivation |
| Identified issues and suggested improvements | Discomfort during prolonged VR use | 7 | Prolonged VR use may cause dizziness |
| | Occasional delays in scene loading | 5 | Loading delays negatively affect user experience |
| | High difficulty level of some inquiry tasks | 4 | Implicit clues increase comprehension time |

The thematic coding analysis results were consistent with the statistical analysis results. The students emphasized immersion, improved comprehension, and motivation as key benefits, while pointing out technical and ergonomic challenges that require further refinement. The learning outcome improvements stemmed from the effects of the event-driven interactive feedback mechanism, the knowledge graph–guided learning mechanism, and the adaptive content generation mechanism based on AICG, rather than from a single instructional factor or incidental intervention.

The analysis results of system operation data, interview feedback, and classroom observations demonstrated the system's effectiveness in supporting learning. However, under high interaction intensity and complex virtual environments, engineering-level optimization is required for further system improvement.

### *4.3. Comparison of LLM and LLM + Knowledge Graph*

To assess the necessity of the Knowledge Graph integration, we conducted a comparative analysis of the standalone ChatGLM-6B model and the proposed KG-augmented architecture, based on established empirical benchmarks for knowledge-intensive tasks [25].

While standalone LLMs demonstrate high linguistic fluency, they exhibit significant reliability problems (hallucinations) when applied to specific historical domains. Comparative studies using benchmarks such as the Benchmark for Fine-grained Automatic Evaluation of Hallucination and MoviE Text Audio Question and Answering show that vanilla LLMs typically show an accuracy rate of 62–68% on domain-specific fact-checking [25]. In contrast, architectures that integrate structured knowledge graphs through retrieval-augmented generation increase the accuracy to higher than 88–92% on the basis of the generative process in verified factual triples [26]. The primary failure modes for the standalone LLMs include temporal conflation and entity misattribution. For example, standalone models frequently integrate the diplomatic contexts of the First and Second Opium Wars or attribute the construction of the Humen Cannon to incorrect historical

figures [26]. These errors are caused by the probabilistic nature of LLMs, which prioritize plausible-sounding sentences over factual precision [27].

The developed system in this study mitigates these errors by injecting verified knowledge triples (e.g., Humen Cannon, Location, and Humen Town) directly into the prompt context. This ensures that the AI's pedagogical responses remain factually based on the museum's data. This approach reduces hallucination rates and enhances the model's ability to handle zero-shot inquiries about local history that were not present in its original training corpus [27].

## 5. Discussion

### *5.1. Teaching Effectiveness and User Experience*

The system is a blended learning tool for instructors. Instructors serve as pedagogical facilitators who frame the historical inquiry before the VR session. Instructors monitor student progress through a centralized dashboard and lead a post-experience synthesis discussion to help students connect the virtual experience to broader historical themes [19]. Regarding accuracy control, while instructors cannot verify every individual AI response in real-time, the knowledge graph acts as an automated symbolic moderator. The system includes a flagging feature: any response generated with a low Bayesian confidence score is sent to the instructor's tablet in real time, allowing the moderator to provide immediate verbal correction or intervention if educational accuracy is compromised [18].

The results of the survey and analysis show that the AI virtual museum significantly enhanced learning outcomes. The experimental group's average score increased from a baseline of 61.2 to 84.5 after the intervention, whereas the control group only reached 71.6. This significant divergence ($p < 0.001$) indicates that the integration of AICG and knowledge graphs facilitates higher-level cognitive mastery compared with traditional methods.

System log data and student feedback indicated that high-intensity concurrent access and complex 3D rendering increased response latency and reduced frame rates. These technical bottlenecks contributed to reports of mild physical discomfort among several participants. This result aligns with established research on cybersickness in virtual reality, which identifies latency and rendering inconsistencies as primary drivers of physiological strain. To address these engineering challenges within the front-end interaction layer, computational loads for non-critical background elements need to be reduced while maintaining high-fidelity responsiveness in core interaction zones. It is also necessary to optimize command-recognition mechanisms to minimize latency and improve immersive quality [28].

While the system was effective for history education, interaction logs revealed a demand for economic and social history content. The modular design of the data and model layer enables seamless content expansion. By leveraging the structured nature of knowledge graphs, the system enriches its node structure without altering its architecture. By refining the artifact–event–knowledge point mapping, new information is naturally embedded into existing interaction paths, preventing the fragmentation of historical knowledge [29].

Post-test analysis and interviews highlighted that immersion alone is insufficient for deep learning. Several students hesitated during open-ended inquiry. This suggests that the core service layer must be improved for an active guide. The integration of AICG through ChatGLM provides dynamic, adaptive content. By utilizing the AICG scheduling module, the system generates staged prompts and visualized learning reports. This transition from static digital repositories to intelligent generative environments follows the global trend in AICG innovation [30]. For further improvement of the AI virtual museum system, standardized task-guidance templates must be added for instructors to effectively organize hierarchical inquiry (Appendix A).

The significance of this system is its ability to facilitate inquiry-based learning, rather than its rendering power. Unlike traditional museums where the narrative is fixed, AICG integration provides dynamic curation, where the history is explored through student-led questions. This transforms the role of a student from a spectator to an active investigator, a transition that is fundamental to developing higher-order historical thinking skills [18].

The primary technical cause of observed latency may stem from the high polygon counts required for 3D artifact reconstruction and the computational overhead of real-time denoising in Stable Diffusion. Long-term solutions are necessary to implement Level of Detail algorithms to reduce rendering complexity for distant objects and to transition to edge rendering to minimize transmission delays between the AI core and VR terminals [31].

### *5.2. AI's Contribution: Static VR and AI-Augmented VR*

To understand the value of the AI components, it is necessary to contrast the system with traditional static VR environments. Static VR, that is, without AI, is used by most virtual museums as digital archives where users follow fixed paths and read pre-written placards [32]. Interaction is limited to basic navigation (roaming). In such environments, the learner is a passive observer, and the depth of inquiry is constrained by the initial script.

In contrast, intelligent VR with AI integrates ChatGLM and knowledge graphs. The developed system in this study transforms visitors' experiences into a dynamic inquiry model. AI enables contextual dialogue, where students ask non-scripted questions and receive historically accurate answers and generative visualization, where stable diffusion is used to create scenes on-demand based on user curiosity. AI also provides adaptive scaffolding, adjusting the complexity of historical relationships based on user interaction logs [33].

Therefore, a considerable difference emerges in the transition from passive consumption to active knowledge construction. While traditional VR provides visual immersion, the AI layer enables cognitive engagement [32]. The 12.9-point score increase in the experimental group in this study is largely attributed to the interactive AI layer, which processes historical causality that static environments cannot offer and is supported by the robust effect sizes observed in our statistical analysis [34].

### *5.3. Limitations*

Despite the results of this study, limitations must be addressed. First, the study was conducted with a relatively small sample of 83 students in a single university in Wuhan, China, which might limit the generalizability of the results to diverse educational institutions. Second, the system's reliance on high-end VR hardware can confine widespread adoption of the developed system in resource-constrained environments. Such limitations necessitate developing lightweight AI models for mobile-based AR and exploring multi-user social interactions in virtual museums. It is also necessary to integrate adaptive difficulty algorithms to adjust inquiry tasks based on real-time student performance logs.

## 6. Conclusions

We developed and validated an AI virtual museum that leverages AICG and knowledge graphs to transform historical education. The system showed its positive effect on students' learning outcomes in Chinese history. The developed system improved student learning outcomes from 61.2 to 84.5 points, with 73.1% reaching the high score level. These results show that the system's integration of intelligent response mechanisms effectively supports complex analytical thinking. The results of this study provide an engineering-level basis for combining large language models (ChatGLM) for dialogue and generative models (Stable Diffusion) for scene reconstruction into a unified, responsive architecture.

The integration of the knowledge graph ensured that AI-generated content maintained historical accuracy, solving a hallucination problem common in generative AI applications. The developed system demonstrated how complex, multi-layered systems can be applied to address domain-specific challenges in digital education. The importance of human-system interaction and the use of real-time behavioral logs was verified to refine system performance.

The results of this study provide preliminary evidence suggesting that the AI virtual museum has a positive effect on students' learning outcomes in Chinese history. During the four-week trial, students using the developed system improved learning outcomes, with 73.1% reaching the high score level. While the findings are promising, the limited sample size and duration suggest that they represent an exploratory validation of the system's potential rather than a definitive measure of its educational impact. In addition, user feedback indicated hardware-induced cybersickness and rendering latencies during high-frequency concurrent access as challenges. It is necessary to optimize the hierarchical rendering pipeline and expand the knowledge graph to include diverse social and economic historical modules, ensuring the system's sustainable scalability and its evolution into a more adaptive, intelligent educational ecosystem.

**Author Contributions:** Conceptualization, Y.L.; methodology, Y.L.; software, S.T.; validation, S.T.; investigation, Y.L.; resources, Y.L.; data curation, S.T.; writing—original draft preparation, S.T.; writing—review and editing, Y.L.; project administration, L.W.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Department of Education of Hubei Province under Project No. 23D047.

**Data Availability Statement:** The datasets generated during this study are fully available within the article.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

The following templates (translated from Chinese) illustrate the structured instructions to standardize task guidance for instructors to effectively organize hierarchical inquiry.

| Template | Purpose | Prompt |
| --- | --- | --- |
| 1 | System identity and safety guardrails | "You are a professional history educator and museum guide for the Virtual Museum of Modern Chinese History. Your goal is to provide accurate, respectful, and pedagogically sound information. If a user asks a question outside of history or the museum's scope, politely redirect them back to the exhibits." |
| 2 | Knowledge-graph-guided inquiry (retrieval-augmented generation) | "Context Information: {KG_Triples}. Based strictly on the provided facts above, explain the historical significance of {Artifact_Name} to the student. Do not include information not supported by the context." |
| 3 | Historical persona roleplay | "Adopt the persona of Lin Zexu in 1839. Speak with the gravity and determination of a Qing official. When asked about the destruction of opium at Humen, describe your motivations and the challenges you faced, maintaining historical fidelity." |

| Template | Purpose | Prompt |
| --- | --- | --- |
| 4 | Socratic pedagogical strategy | "Instead of providing a direct answer to the student's question about {Event}, ask a leading question that encourages them to analyze the cause-and-effect relationship based on the museum's timeline." |
| 5 | Conflict resolution (Bayesian output) | "The user believes {User_Claim}, but our knowledge base indicates {KG_Fact}. Acknowledge the user's perspective but gently correct the record using the provided evidence, citing the specific display in the virtual hall." |

## References


1. Foo, L.G.; Rahmani, H.; Liu, J. AI-generated content (AICG) for various data modalities: A survey. *ACM Comput. Surv.* **2025**, *57*, 1–66. [CrossRef]
2. Wang, J. Research on the design and implementation strategy of personalized art education experience based on AICG. *SHS Web Conf.* **2025**, *213*, 1–8. [CrossRef]
3. AlBlooish, S. Artificial intelligence in history education: Opportunities and challenges. *Fron. Educ.* **2026**, *29*, 1683968. [CrossRef]
4. Gherardi, E.; Benedetto, L.; Matera, M.; Buttery, P. Using Knowledge Graphs to Improve Question Difficulty Estimation from Text. *Lect. Notes Comput. Sci.* **2024**, *14830*, 293–301. [CrossRef]
5. Rodriguez-Garcia, B.; Guillen-Sanz, H.; Checa, D.; Bustillo, A. A systematic review of virtual 3D reconstructions of Cultural Heritage in immersive Virtual Reality. *Multimed. Tools Appl.* **2024**, *83*, 89743–89793. [CrossRef]
6. Jimeno, A.; Puerta, A. State of the art of the virtual reality applied to design and manufacturing processes. *Int. J. Adv. Manuf. Technol.* **2007**, *33*, 866–874. [CrossRef]
7. Zou, Y.; Awad, A.; Lin, M. DirectNVM: Hardware-accelerated NVMe SSDs for high-performance embedded computing. *ACM Trans. Embed. Comput. Syst.* **2022**, *21*, 1–24. [CrossRef]
8. Hijma, P.; Heldens, S.; Sclocco, A.; van Werkhoven, B.; Bal, H.E. Optimization techniques for GPU programming. *ACM Comput. Surv.* **2023**, *55*, 1–81. [CrossRef]
9. Chickerur, S.; Balannavar, S.; Hongekar, P.; Prerna, A.; Jituri, S. WebGL vs. WebGPU: A performance analysis for Web 3.0. *Procedia Comput. Sci.* **2024**, *233*, 919–928. [CrossRef]
10. Li, M.; Zhao, Y.; Yu, B.; Song, F.; Li, H.; Yu, H.; Li, Z.; Huang, F.; Li, Y. API-Bank: A comprehensive benchmark for tool-augmented LLMs. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023. [CrossRef]
11. Davuluri, P.S.L. (Ed.) Cloud-Native Architectures for Scalable Data Systems. In *The Autonomous Data Enterprise: Engineering Real-Time Intelligence with Generative and Agentic AI*; DeepScience: San Francisco, CA, USA, 2026; pp. 34–51. [CrossRef]
12. Han, C. Application of AICG image generation technology in product design. *Adv. Comput. Mater. Sci. Res.* **2025**, *1*, 161–164. [CrossRef]
13. Kim, T.W. Application of artificial intelligence chatbots, including ChatGPT, in education, scholarly work, programming, and content generation and its prospects: A narrative review. *J. Educ. Eval. Health Prof.* **2023**, *20*, 38. [CrossRef]
14. The Supreme People's Procuratorate of the People' Republic of China. Available online: https://en.spp.gov.cn/introduction.html (accessed on 27 February 2026).
15. Zeng, A.; Xu, B.; Wang, B.; Zhang, C.; Yin, D.; Zhang, D.; Rojas, D.; Feng, G.; Zhao, H.; Lai, H.; et al. ChatGLM: A family of large language models from GLM-130B to GLM-4 all tools. *arXiv* **2024**, arXiv:2406.12793. [CrossRef]
16. Liu, B.; Zhao, W.; Wang, J.; Yan, J.; Peng, J. Construction and application of multi-modal knowledge graph for collection of cultural relics. In Proceedings of 2nd International Conference on Image Processing and Media Computing, Xi'an, China, 26–28 May 2023. [CrossRef]
17. Tserklevych, V.; Prokopenko, O.; Goncharova, O.; Horbenko, I.; Fedorenko, O.; Romanyuk, Y. Virtual museum space as the innovative tool for the student research practice. *Int. J. Emerg. Technol. Learn.* **2021**, *16*, 213–231. [CrossRef]
18. Pearl, J. Causal diagrams for empirical research. *Biometrika* **1995**, *82*, 669–688. [CrossRef]
19. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T.; et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In Proceedings of the 34th Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2020; pp. 9459–9474. Available online: https://dl.acm.org/doi/abs/10.5555/3495724.3496517 (accessed on 11 March 2026).

20. Florou, K. Using NLP Tools to Enhance Italian Language Teaching. In Proceedings of the PIXEL Conference: The Future of Education 13th, Firenze, Italy, 29–30 June 2023. Available online: https://www.researchgate.net/publication/385346158_Using_NLP_Tools_to_Enhance_Italian_Language_Teaching_A_Qualitative_Study_in_Higher_Education (accessed on 11 March 2026).
21. Li, F.; Gao, Y.; Candeias, A.J.E.G.; Wu, Y. Virtual Restoration System for 3D Digital Cultural Relics Based on a Fuzzy Logic Algorithm. *Systems* **2023**, *11*, 374. [CrossRef]
22. Chen, Y.; Bao, J.; Weng, G.; Shang, Y.; Liu, C.; Jiang, B. AI-enabled multi-mode electronic information innovation practice teaching reform prediction and exploration in application-oriented universities. *Systems* **2024**, *12*, 442. [CrossRef]
23. Zhang, S. Character recognition of historical and cultural relics based on digital image processing. In Proceedings of 5th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 2–4 December 2021. [CrossRef]
24. Shapiro, B.J. The subjective estimation of relative word frequency. *J. Verb. Learn. Verb. Behav.* **1969**, *8*, 248–251. [CrossRef]
25. Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; Wu, X. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 3580–3599. [CrossRef]
26. Wang, Y.; Cai, Y.; Chen, M.; Liang, Y.; Hooi, B. Primacy Effect of ChatGPT. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, Singapore, 6–10 December 2023; pp. 120–135. [CrossRef]
27. Zhu, Y.; Wang, X.; Chen, J.; Qiao, S.; Ou, Y.; Yao, Y.; Deng, S.; Chen, H.; Zhang, N. LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities. *World Wide Web* **2024**, *27*, 58. [CrossRef]
28. Rebenitsch, L.; Quebbeman, S. Review on cybersickness in applications and visual displays. *Virtual Real.* **2016**, *20*, 101–125. [CrossRef]
29. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 494–514. [CrossRef]
30. Cao, Y.; Li, S.; Liu, Y.; Yan, Z.; Dai, P.; Yu, P.S.; Sun, L. A comprehensive survey of AI-generated content (AICG): A history of generative AI from GAN to ChatGPT. *arXiv* **2023**, arXiv:2303.04226. [CrossRef]
31. Chang, E.; Kim, H.T.; Yoo, B. Virtual Reality Sickness: A Review of Causes and Measurements. *Int. J. Hum.–Comput. Interact.* **2020**, *36*, 1658–1682. [CrossRef]
32. Radianti, J.; Majchrzak, T.A.; Fromm, J.; Wohlgenannt, I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Comput. Educ.* **2020**, *147*, 103778. [CrossRef]
33. Wang, F.; Zhou, X.; Li, K.; Cheung, A.C.K.; Tian, M. The effects of artificial intelligence-based interactive scaffolding on secondary students' speaking performance, goal setting, self-evaluation, and motivation in informal digital learning of English. *Interact. Learn. Environ.* **2025**, *33*, 4633–4652. [CrossRef]
34. Bujang, M.A. A Power Primer Revisited. *Indian J. Psychol. Med.* **2026**, *in press*. [CrossRef]


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.