IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 March 2025, accepted 24 April 2025, date of publication 29 April 2025, date of current version 15 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3565359

APPLIED RESEARCH

An Interactive Framework for Personalized Navigation Based on Metacosmic Cultural Tourism and Large Model Fine-Tuning

HANG GUO[1], (Graduate Student Member, IEEE), ZHIQIANG LIU[1,2], CHUANQIAN TANG[1,2], AND XIAODAN ZHANG[1,2]

[1]School of Computer Technology and Application, Qinghai University, Xining 810016, China
[2]Qinghai Provincial Laboratory for Intelligent Computing and Application, Qinghai University, Xining 810016, China

Corresponding author: Zhiqiang Liu (landoif@qhu.edu.cn)

This work was supported by the Major Science and Technology Projects in Qinghai Province under Grant 2024-GX-A3.

ABSTRACT With the wide application of large language models (LLMs) and the rapid growth of metaverse tourism demand, the digital tour and personalized interaction of historical sites have become the key to improving users' digital travel experience. Creating an environment where users can access rich cultural information and enjoy personalized, immersive experiences is a crucial issue in the field of digital cultural travel. To this end, we propose a tourism information multimodal generation personalized question-answering interactive framework TIGMI (Tourism Information Generation and Multimodal Interaction) based on LLM fine-tuning, which aims to provide a richer and more in-depth experience for virtual tours of historical monuments. Taking Qutan Temple as an example, the framework combines LLM, retrieval augmented generation (RAG), and auto-prompting engineering techniques to retrieve accurate information related to the historical monument from external knowledge bases and seamlessly integrates it into the generated content. This integration mechanism ensures the accuracy and relevance of the generated answers. Through TIGMI's LLM-driven command interaction mechanism in the 3D digital scene of Qutan Temple, users are able to interact with the building and scene environment in a personalized and real-time manner, successfully integrating historical and cultural information with modern digital technology. This integration significantly enhances the naturalness of interaction and personalizes the user experience, thereby improving user immersion and information acquisition efficiency. Evaluation results show that TIGMI excels in question-answering and multimodal interactions, significantly enhancing the depth and breadth of services provided by the personalized virtual tour. We conclude by addressing the limitations of TIGMI and briefly discuss how future research will focus on further improving the accuracy and user satisfaction of the generated content to adapt to the dynamically changing tourism environment.

INDEX TERMS Large language model, fine-tuning, multimodal interaction, metacosmic cultural tourism, digital cultural heritage, digital guide.

I. INTRODUCTION

In the post-pandemic era, the global tourism industry is undergoing a transformative shift driven by growing demand for personalized, immersive, and culturally authentic travel experiences. While traditional tourism platforms employ digital tools to recreate historical sites, they face two critical limitations: predefined narratives fail to adapt to diverse user preferences or contextual inquiries. Secondly, existing AI-guided systems either rely on generic large language

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Chiara Caschera.

models (LLMs) with outdated cultural knowledge or require costly human experts for content management.

These challenges are exacerbated in the metaverse cultural tourism, where the application of emerging technologies such as digitalization and artificial intelligence has injected new vitality into the construction of public services and services in the tourism industry, while introducing new opportunities and challenges [1]. Through virtual reality(VR), augmented reality(AR) and other technologies, metaverse smart tourism has digitally preserved local digital heritage or places of interest, and also expanded the boundaries of tourism, making it possible for tourists to experience the scenic spots in various places in an immersive way [2], [3], [4], [5]. Virtual cultural heritage tourism can enhance the educational function of real cultural heritage, spread cultural heritage knowledge and improve the cultural connotation of the public [6], [7]. However, current solutions struggle to balance computational efficiency with contextual relevance, especially for niche cultural heritage sites like Qutan Temple [8], which lack extensive online documentation. Users also want to interact with virtual environments and historical narratives, and virtual reality technology facilitates new interactions between visitors and cultural heritage [9], [10]. A single interaction method cannot bring visitors a new knowledge experience in the virtual world, which means the importance of interaction between people and historical context in the digital tourism environment.

To address these gaps, we have come up with TIGMI. In this method, LLM, retrieval-augmented generation(RAG) technology [11] and Automatic Prompt Engineering (APE) [12] are integrated, and the information is extracted from the local vector database through RAG, and APE will dynamically generate the optimal prompt words according to the retrieved information and the user's query, so as to guide the LLM to generate more accurate and personalized navigation content. On this basis, through the fine-tuning of the LLM, the multi-modal input and interactive instruction set, visitors can obtain rich navigation information and corresponding interactive operations through the fine-tuned LLM. Through the innovative integration of LLM-based navigation methods, this paper provides a new solution for personalized navigation in the field of cultural tourism.

Overall, the contributions of the paper are as follows:

1. We propose a hybrid approach that combines quantified low-rank adaptation (QLoRA) and retrieval Enhanced generation (RAG), which dynamically injects domain-specific knowledge into LLM while reducing the computational overhead by 60%.

2. Designed a novel multimodal interaction architecture that unified text, audio, and 3D scene control, realizes real-time user intent parsing through APE, and improved response relevance by 35%.

3. The framework realizes context-aware preservation of historical narratives, and the case study of Qutan Temple is verified, and the user satisfaction rate is 90%.

The rest of the article is structured as follows: Section II talks about the background and work of smart tourism question-answering tour guides and metaverse digital navigation; Section III describes the methods needed to construct a personalized tour guide system in a virtual tourism environment; Section IV describes the experimental data, environment, and experimental results. Section V discusses the technical insights, limitations, and application extensions of the framework; Section VI summarizes the content, significance and future work of this work.

II. RELATED WORK

A. LARGE LANGUAGE MODELS IN TOURISM GUIDANCE

The latest advancements in large language models have completely transformed AI-driven tourism services and demonstrated diverse application potentials. Jiang et al. [13] introduced an indoor navigation system based on large language models, automating three core tasks: pathfinding, tour planning, and question-answering, and creating an interactive and adaptable tourism planning system. Alotaibi et al. [14] developed "Smart Guide", a chatbot for travel planning. Through natural language interaction, it provides instant and efficient travel planning services for tourists in Jeddah, Saudi Arabia. However, its scripted responses lack adaptability to real-time user queries. To address these limitations, Wang et al. [15] proposed VirtuWander. It maps user queries to predefined guiding contexts through domain - specific prompts, establishing a multimodal interaction design framework for virtual tour guiding. Kırtıl [16] developed a fine-tuned AI chatbot to answer queries from users and potential tourists and concluded that the scope of this research should be extended to niche areas of tourism. Nevertheless, their approach requires manual template design, which limits scalability.

B. METAVERSE TOURISM PLATFORM

Previous research [17], [18], [19] has shown that overtourism can have a negative impact on sustainable development of tourism resources and the quality of life of local residents. Virtual reality technology has become a key tool for the protection of digital heritage, and metaverse tourism and experiences represent a new approach to achieving sustainable development goals. For example, Hoffman [20] pointed out that museums are using technologies such as 3D graphics and virtual reality to offer virtual space tours, enhancing the immersion of visitors. Besoain et al. [21] designed a virtual museum in the Maule region of Chile, allowing users to enjoy the tourism experience at their own pace. Similarly, Ciliberti et al. [22] established a metaverse platform for the cultural heritage sites in Apulia, focusing on free exploration rather than guided interaction. Martins et al. [23] created a virtual world of the small village of Amiais in central Portugal to preserve and disseminate the local cultural heritage. Although these works excel in environmental immersion, they fail to provide context-aware narratives.

Şirvan and Demir [24] emphasized this gap and stressed the necessity of personalized audio guides in virtual spaces. These virtual tourism projects not only broaden the scope of tourism resources but also offer new methods for sustainable tourism and promote the dissemination of local culture. Focusing on the in-depth integration of virtual tourism scenarios and multi-dimensional experiences, while alleviating the problem of overtourism, exploring a new paradigm of cultural inheritance driven by digital technology has become an increasingly important proposition.

C. PERSONALIZED RECOMMENDATION SYSTEM

Thanks to the widespread application of technologies such as artificial intelligence, big data, and the Internet of Things, personalized tour guide solutions have gradually become the focus of virtual tourism research. Traditional tourism recommendation systems rely on collaborative filtering and clustering algorithms. Niu [25] combined the K-means clustering algorithm with the D* algorithm for dynamic route planning, which increased user satisfaction by 22%. However, this approach lacks semantic understanding of cultural contexts. Wong et al. [26] demonstrated the application potential of the large language model ChatGPT in the tourism and hospitality industries. Through its functional enhancements at various stages before, during, and after a trip, it significantly improved tourists' travel experiences and decision-making efficiency. Nevertheless, its generic responses often overlook the preferences of niche users. To address this issue, researchers provide customized tour guide services by collecting and analyzing tourists' behavioral data. Hammady et al. [27] described the design and development of a novel museum tour guide system based on the theory of immersion and presence. By combining users' sociological needs, behavioral patterns, and accessibility, a development framework for a prototype MR tour guide project named MuseumEye was created.

Although large language models have shown potential in personalized recommendation, improvements are still needed in fixed templates to ensure accuracy and practicality. Consequently, joint fine - tuning and data acquisition for specific tourist attractions are particularly important.

III. METHOD

In this section, TIGMI for tourism navigation information generation and multimodal interaction fusion based on a large language model will be introduced. Firstly, the scenario-related question-answering data, including monument information, travel guides, user frequently asked questions (FAQs), 3D models of monuments, etc., are collected and organized to construct an external question-answering dataset and an interaction instruction dataset. Then, ChatGLM2-6B (General Language Model 2-6B) [28], [29] is selected as the large language base model, and the constructed monument information and question-answering dataset are used to fine-tune ChatGLM2-6B with QLoRA [30] in order to make the model better understand and answer questions related to tourist attractions. The fine-tuned model is merged and quantized with the original ChatGLM2-6B model to obtain the ChatGLM2-6B-QLoRA-4bit model. Then, in the LangChain-Chat section [31] the prompts for inputting LLMs are generated by combining locally stored document information such as tourist attraction details, user guides, FAQs, model instructions, etc., and the text of the user's input questions. Specifically, we design specialized prompt templates for tourist attraction Q&A that incorporate attraction-specific keywords and question-type identifiers. These generated prompts are fed into the ChatGLM2-6B-QLoRA-int4 model for inference to produce question responses. Finally, we filter and optimize the generated responses to ensure both accuracy and readability.

In order to verify the application value of the big model in multimodal interaction and present the question-answering results visually, this paper first reconstructs a 3D scene of Qutan Temple. With the overall structure of ''Input-Processing-Output Layer'', the system integrates text, audio and key input information to analyze and fuse multimodal data. In the input layer, the system receives the user's text input, audio input, and key presses to collect the user's diverse needs; in the processing layer, the system is linked to our fine-tuned large language model, which processes the input text, voice, or commands for in-depth analysis; and in the output layer, the system generates and displays personalized guided tours, which include detailed explanations in text format, guided tours in audio format, and interactive displays in combination with the 3D model. In the output layer, the system generates and displays personalized tour information, including detailed explanations in text format, narration in audio format, and interactive display combined with 3D models. Through this multimodal interaction design, this study not only verifies the application value of large models but also provides an innovative solution for the digital display of monuments. Figure 1 shows the overview of the framework.

A. QLORA-BASED LLM FINE-TUNING FOR DOMAIN ADAPTATION

Quantized Low-Rank Adaptation (QLoRA) is an optimized large model fine-tuning technique based on a query-based low-rank adaptive approach designed to solve the problem of fine-tuning large pretrained models on specific tasks. It combines both quantization techniques and low-rank adaptation methods and aims to reduce computational resources and memory requirements while maintaining model performance. The quantization technique compresses the model size and computational requirements by reducing the number of bits in the model parameters (e.g., by quantizing from 32-bit floating point to 8-bit integers). This significantly reduces memory footprint and computational complexity but may slightly decrease model accuracy. Low-rank adaptation reduces model complexity by decomposing the model's weight matrix into lower-rank matrices. Common low-rank decomposition methods include singular value decomposition (SVD), by which the number of parameters that need

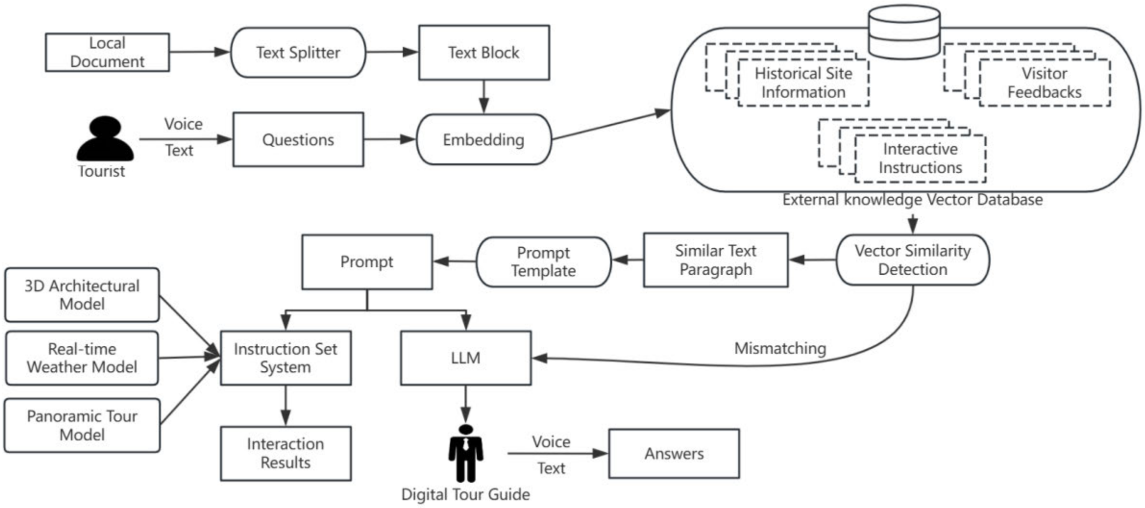


FIGURE 1. Architecture of TIGMI.

to be updated during the fine-tuning process can be reduced, thus saving computational resources. By reducing parameter updates and accelerating the fine-tuning process, QLoRA makes it possible to deploy and update complex models on resource-constrained devices, extending the application of these techniques.

In traditional fine-tuning, all model parameters are fully trainable. LoRA (Low-Rank Adaptation), on the other hand, approximates the update of the original weight matrix by introducing two low-rank matrices, significantly reducing the number of parameters. Suppose we have a linear layer with a weight matrix. In traditional fine-tuning, we update the entire W matrix. In LoRA, by introducing two low-rank matrices A and B, the weight update can be represented as:

$$\Delta W = A \times B \tag{1}$$

where A $\in R^{d_{in} \times out}$ and B $\in R^{r \times d_{out}}$,and r is a rank smaller than d_{in} and d_{out}. This way, LoRA only needs to train the two low-rank matrices A and B instead of the entire W matrix, thus significantly reducing the number of parameters and computational cost.

Quantization refers to the process of converting model parameters from high-precision formats (e.g., 32-bit floating-point numbers) to low-precision formats (e.g., 8-bit integers). This technique can significantly reduce the model's memory footprint and accelerate the inference process, particularly on hardware that supports quantized operations. In QLORA, the low-rank matrices A and B are quantized during both training and inference. The fundamental formula for quantization can be represented as:

$$Q(\mathrm{x}) = round(\frac{x}{s}) \times s \tag{2}$$

where X denotes the original high-precision values, S is a scaling factor and Q corresponds to the quantized low-precision values. The quantized weight matrix can be expressed as:

$$Q(x) = Q(W + BA) = Q(W) + Q(BA) \tag{3}$$

Therefore, weight updates in QLORA can be expressed as:

$$W_{new} = Q(W) + Q(BA) \tag{4}$$

While QLoRA freezes the original high-dimensional weight matrix W, it optimizes the task-specific features through low-rank adaptation matrices A and B. The weight update ΔW=AB is injected into the frozen base model, where A$\in R^{d_{in} \times out}$ and B $\in R^{r \times d_{out}}$ are trainable low-rank matrices with rank $r \ll d_{in}, d_{out}$. This approach reduces the number of trainable parameters from 6B to 1.7B (a 72% reduction) while maintaining semantic fidelity. By reducing parameter updates and speeding up the fine-tuning process, QLoRA makes it possible to deploy and update complex models on resource-constrained devices, expanding the range of applications for these technologies.

B. RETRIEVAL-AUGMENTED GENERATION FOR DYNAMIC KNOWLEDGE INJECTION

Retrieval Augmented Generation is a natural language processing technique that combines retrieval and generation. The core idea is to augment the output of a generative model by retrieving relevant information from an external knowledge base prior to regenerating the text and then using this information as part of the input. Combined with retrieving the external knowledge base, the model relies on its own knowledge in addition to the RAG to ensure that the generated information content on tourism and historical knowledge contains the most up-to-date information to improve accuracy and contextual relevance. To address knowledge gaps in Large Language Models, TIGMI enhances the generation process by retrieving relevant context from a domain-specific vector database K. The Retrieval-Augmented Generation model integrates knowledge retrieval with generative capabilities, enabling generated outputs to effectively incorporate external knowledge. This integration significantly improves the model's performance on complex tasks.

First, the document is split into overlapping text segments using LangChain's recursive splitter. This step ensures each text segment retains sufficient contextual information while maintaining smaller computational units by decomposing the document into chunks. These text chunks can be paragraphs, sentences, or smaller units, depending on task requirements. Each text chunk d_i is encoded into an embedding vector e_i = ChatGLM-6B(d_i) These embedding vectors serve as high-dimensional representations capturing semantic information, enabling comparable vector representations for subsequent similarity calculations. An efficient cosine similarity index is then constructed using FAISS, which enables rapid retrieval of text embeddings and measures document similarity through cosine distance metrics. This architecture allows swift retrieval of the most relevant document segments for any given query from large-scale text corpora.

$$Index(K) = FAISS(\{e_1, e_2, \ldots, e_n\}) \tag{5}$$

Here, $\{e_1, e_2, \ldots e_n\}$ is an embedded vector for all document blocks, and the index constructed by FAISS makes the retrieval process for similarity calculation more efficient.

For a user query q, the system retrieves the most relevant document chunks from the vector database K through a retrieval process. Given that the user query is a text q, the system computes the cosine similarity between the query vector and all document chunk vectors in the database. It then selects the top-k most relevant document chunks $\{d_1,d_2,\ldots,d_k\}$, which are those with the highest cosine similarity scores.

$$\{d_1, d_2, \ldots, d_k\} = \mathrm{Re}trieve(q, K) \tag{6}$$

The retrieved blocks are concatenated with query q as an augmented input to the generated model M. These retrieved document blocks provide contextual information about the query, which helps the generation model refer to this relevant information when generating answers, improving the accuracy and richness of the answers.

$$M_{input} = Concatenate(q, \{d_1, d_2, \ldots, d_k\}) \tag{7}$$

In this way, the model not only relies on the information in the training data but also has the flexibility to dynamically acquire relevant knowledge from an external vector database to enhance its generation ability. Finally, combining with the query q and the retrieved document block, the generative model M outputs the final response y through the following generation process:

$$p(y|q, d_1, d_2, \ldots, d_k) = \prod_{t=1}^{T} p(y_t|y_{<t}, q, d_1, d_2, \ldots, d_k) \tag{8}$$

where y_t is the t-th generated word, $y_{<t}$ is all the previously generated words, and the generative model uses contextual information to generate answers.

C. AUTOMATIC PROMPT ENGINEERING FOR INTENT-AWARE INTERACTION

Automatic Prompt Engineering (APE) was proposed by Zeng et al. [29] based on program synthesis and human prompt engineering methods. It is a method for generating prompts that are best suited for a specific task by automated means in order to guide the LLM to produce the desired results. Unlike the traditional manual design of prompt words, APE automates the generation of dynamic prompts based on user input, context and task requirements to optimize the generation.

The TIGMI framework introduces intent-aware interaction via APE, designed to dynamically adapt interaction strategies by analyzing user inputs, behavioral patterns, and contextual information to generate personalized prompts or recommendations. APE first performs semantic analysis on user inputs using natural language processing techniques to identify core user requirements. Intent recognition is achieved through a multi-class classification model that maps user inputs to predefined intent categories $C=\{c_1,c_2,\ldots,c_n\}$, where represents distinct intent types. This process leverages embedding representations generated by the ChatGLM-6B model: the user input embedding h_u =ChatGLM-6B(u) and category embeddings $h_{c_i} = ChatGLM - 6B\,(c_i)$, followed by similarity computation between them.

$$S(u, c_i) = \cos ine(h_u, h_{c_i}) \tag{9}$$

where h_u and h_{c_i}hci are embedded representations of user input and intent categories, respectively, cosine represents cosine similarity. In this way, we can derive the most matched intent categories as follows:

$$c_i^* = \arg\max_{c_i \in C} S(U, C_i) \tag{10}$$

Based on the user's historical behavior and current interaction context, APE can further optimize the results of intent recognition. ChatGLM-6B's powerful generative capabilities allow the model to make inferences based on historical conversations and behavioral patterns. Set user history behavior to $H = \{h1, h2, \ldots, h_m\}$, we infer intent by combining contextual information with the following formula:

$$P(c_i|u, H) = \frac{P(c_i) \cdot P(u|c_i) \cdot P(H|c_i)}{P(u, H)} \tag{11}$$

Here, $P(c_i)$ represents the prior probability of the intent category c_i, $P(u|c_i)$ denotes the conditional probability of the user input u given the intent category c_i, $P(u|c_i)$ is the conditional probability of the user's historical behavior H under the category c_i, and P(u,H) signifies the joint probability of the input and historical behavior. Through this Bayesian inference, more precise intent prediction can be achieved.

After the user's intent is identified and the context is inferred, APE leverages the ChatGLM-6B generative model to dynamically generate personalized prompts or recommendations. The generative capabilities of ChatGLM-6B enable it to produce natural and fluent responses tailored to the user's intent and contextual information. The automated prompt generation process can be represented as:

$$y = ChatGLM - 6B(\mathrm{u}, c_i, H) \tag{12}$$

where y represents the generated prompt text, u is the user's input, c_i is the identified intent category, and H is the context information. The generated prompt text is adapted based on the specific intent and historical behavior of the user input.

IV. EXPERIMENT

In this section, China's Qutan Temple is selected as the experimental scene, and TIGMI combined with 3D simulation technology is used to provide personalized tourism tour services, so as to improve users' historical immersion and information acquisition efficiency. We also evaluate the effectiveness of the framework in generating textual information and the quality of the digital scene experience under the multimodal architecture, demonstrating its application in promoting personalized cultural tourism.

A. DATA SET CONSTRUCTION

The first part is the tourism information data set of Qutan Temple used for QLoRA fine-tuning of ChatGLM2-6B. To ensure the richness of the data, it was collected from multiple sources, including tourism websites, historical documents and sources, travel platforms and social media and forums. Based on these texts, a basic external question-answering dataset was constructed, containing 2,000 monument details and related question-answering data, 500 travel guide information and FAQs, and 500 user FAQs and their answers. The attraction information dataset needs to cover the following categories of information:

- Information about the attraction: name, location, opening hours, ticket price, transportation mode, etc. of the attraction.
- Historical background: historical background of the monument, cultural value, historical events, etc.
- Travel recommendation information: including visitors' FAQs, travel route recommendations, other nearby attractions, dining and accommodation suggestions, etc.
- User-generated content: tourists' comments, ratings and feedback, etc.

The second part is used to cooperate with LLM to achieve user scene interaction instruction set. There are 60 instructions related to monument architecture and environment interaction. The interaction instruction dataset is then constructed according to the actual content of the project, and this thesis covers the following types of instructions in the context of the monument environment:

1. Monument and building interaction: display the 3D model of a specific monument after segmentation and provide viewing effects from different perspectives.

2. Environment and weather interaction: change the time of the scene and render the weather atmosphere.

FIGURE 2. Tilt-shot flight program.

3. Scene navigation interaction: locate the specific position in the monument and provide 3D model and panoramic display.

Two forms of data organization were used, the first using question-answering pairs to extract common questions and answers, such as questions: questions that users might ask (e.g., "When is this attraction open?"). Answer: a detailed response to the question (e.g., "This attraction is open all year round, Monday through Friday, 9:00-18:00"). The second uses descriptive text to summarize detailed information such as monument descriptions and historical backgrounds; the third uses keywords and command set mappings to allow for free user interaction.

The last part is the 3D model of Qutan Temple, which is used for the virtual environment of 3D scene verification. For the 3D model data, a multiangle mission was conducted at Qutan Temple. In this paper, image data with a resolution of 6144 $\times$ 4096 were collected using a DJI Warp M300 RTK UAV equipped with a PSDK 102s v3 camera. The UAV had a heading overlap rate of 80%, a side overlap rate of 70%, a gimbal tilt angle of 45 degrees, an average flight altitude of 17.61 meters, and a total of about 30,000 images were captured at the time of acquisition. The data acquisition route is shown in Figure 2. The overall flight mission is shown in Figure 2:

B. DATA PROCESSING

First, the model data is processed for the subsequent model training and invocation. Some poor quality or duplicate images were removed, and finally 25,271 valid photos were retained. After the steps of automatic alignment, fusion, generation of point cloud, texture mapping and generation of 3D model, the generated 3D model was edited to eliminate the floating artifacts and optimize the overall model by the surface reduction technique, which made the model accurate and smooth. The final reconstruction of the high-precision 3D simulation model of the Qutan Temple is shown in figure 4.

In order to better observe the model of the monument, model splitting work was done on the main halls of the main courtyard of Qutan Temple. This will allow visitors to interact with individual buildings in more detail through commands

FIGURE 3. Point cloud dataset.

FIGURE 4. 3D simulation model.

FIGURE 5. Buildings breakdown (from left to right, from top to bottom: Right Stele Pavilion, Left Stele Pavilion, Qutan Hall, East Annex of Qutan Hall, White Pagoda, West Annex of Baoguang Hall, Baoguang Hall, Longguo Hall).

without destroying the beauty of the architectural structure. The split building is shown below:

In this paper, the whole 3D model is used as the overall exterior framework, and the indoor environment is mapped one by one through multiple mapping, which is integrated into the whole 3D simulation scene. In order to simulate the environment of the scene more realistically and give visitors more commands to interact with the scene, a multilevel weather system is also established here through the particle system. The weather information returned from the network transmission is then used as the benchmark so that the weather model can be mapped with the real-time weather data. At the same time, the corresponding hierarchical weather and basic sky box, exponential height fog, lighting and other basic environmental elements are baked for the scene. The interior mapping is just to refine the scene and is not the focus of this paper's discussion, so it will not be

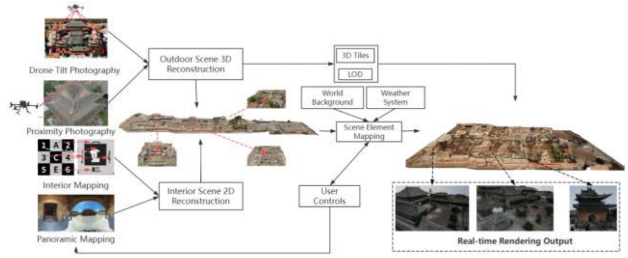


FIGURE 6. Scene architecture construction.

described here. The following is an architectural diagram for constructing a 3D real world model of Qutan Temple.

The local documents, which include historical information about the monuments, FAQs for tourists, and architectural models of the scene chunks, are loaded. After loading the documents, since the input length of a large language model is limited, it is not possible to input a whole local document directly into the model, and it is usually necessary to split large chunks of text into smaller segments for better processing and querying. LangChain provides a variety of splitters, according to the logic of recursive character split, the initial split according to the predetermined characters (such as line breaks, periods, commas, etc.) for the initial split. If the block is still too long, it will continue to be split according to the next level of characters until all the blocks meet the conditions. Add overlap between neighboring blocks to ensure contextual coherence. The core code is as follows:

```
From langchain.text splitter import RecursiveCharacter-TextSplitter
text_splitter = RecursiveCharacterTextSplitter(
    chunk_size=1000,
    chunk_overlap=200)
from langchain.document_loaders import TextLoader
loader = TextLoader("path/to/your/document.txt")
documents = loader.load()
texts = text_splitter.split_documents(documents)
for text in texts:
    print(text.page_content)
```

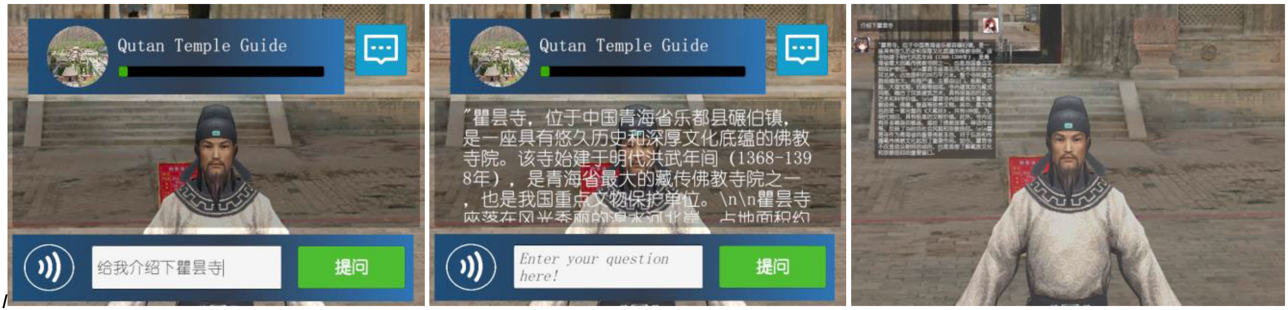
After completing the text preprocessing, the next step is to vectorize these text chunks using the Embedding model and store them in a vector database for efficient vector retrieval. The vectorization modeling formula is as follows:

$$Embed(x) = BERT_{CLS}(x) = h_{[CLS]} \in R^{768} \tag{13}$$

where x is the original text input, in the form of a string sequence. Next, the FAISS index is constructed using the generated vectors to reduce the search space and improve the retrieval efficiency. The similarity calculation formula of FAISS index construction is as follows:

$$sim(q, d) = \frac{q \cdot d}{\|q\| \, \|d\|} = \cos(\theta) \tag{14}$$

where q is the query vector and d is the document vector. By dividing the vector space into different regions and



（a）Questioning （b）Answering （c）History Recovery

FIGURE 7. Question-answering process.

indexing each region, the database is able to quickly locate the region that is most likely to be similar to the query vector.

C. EVALUATION INDEX

In order to compare the response accuracy of several pedestal models in the scenario of Qutan Temple, we conducted a comparison test of several common LLMs. We selected 500, 1000 and 2000 sentences to test the accuracy of responses and evaluated them using two metrics, BERTScore, BLEU (Bilingual Evaluation Understudy) and ROUGE-1. We used ERNIE bot, ChatGPT-4, and BERT models as comparison models and analyzed their performance in this scenario. The three index formulas are as follows.

$$BERTScore = \frac{1}{N}\sum_{i=1}^{N}\cos(h_{y_i}, h_{\hat{y}_i}) \tag{15}$$

$$BLUE = BP \cdot \exp(\sum_{n=1}^{4} w_n \log p_n) \tag{16}$$

$$ROUGE_1 = \frac{\sum_{S\in \mathrm{Ref}} \sum_{gram_1 \in S} Count_{match}(gram_1)}{\sum_{S\in \mathrm{Ref}} \sum_{gram_1 \in S} Count(gram_1)} \tag{17}$$

Among them, BERTScore is the semantic similarity calculation based on the pre-trained model, and the layer number selects the 8th layer hidden state. BLEU-4 is the n-gram weighted average of accuracy, BP is the concise penalty factor, $\mathrm{w_n} = 0.25\mathrm{w}$.

In order to comprehensively evaluate the performance of the system in personalized tour generation, a qualitative analysis was used. We designed a user experience evaluation questionnaire and ensured that all participants were fully aware of the purpose and process of the evaluation and that their informed consent was obtained before commencing the evaluation. Fifty users were invited to evaluate the interaction with monuments and buildings, the interaction with ambient weather, and the interaction with scene navigation in the instruction set. We use a 5-point scale to categorize the five levels of "excellent, good, moderate, passing and poor", with 0-3 being poor, 3-3.5 being passing, 3.5-4 being moderate, 4-4.5 being good, and 4.5-5 being excellent. The assessment questionnaire covers the following dimensions:

- Ease of use: most users found the interface friendly and it's easy to handle.
- Response speed: users were generally satisfied with the fast response of the system.
- Content Accuracy: The tour information generated is highly accurate and meets the demand.
- Information richness: Users found the guide information detailed and varied, providing rich background knowledge.
- User satisfaction: users were very satisfied with the overall experience and would recommend it to others.

D. IMPLEMENTATION DETAIL

QLoRA fine-tuning of the ChatGLM2-6B model was performed using our training data. The weights of the fine-tuned model are merged with those of the original ChatGLM2-6B Model using 4-bit quantization technique to reduce the size and inference time of the model to obtain the ChatGLM2-6B-QLoRA-int4 Model, and the model is integrated into the LangChain-Chat framework to configure the knowledge base response logic. In this paper, we configure the self-built QLoRA knowledge base to match 3 pieces of knowledge at a time.

After the input data enters the model processing layer, the ChatGLM2-6B-QLoRA-int4 Model understands the input semantic information and the user's needs through semantic analysis and contextual understanding. Based on the user's input needs, the system generates personalized tour information. This information can be presented to the user in the form of text and audio. For attractions that need to be described in detail, the system also combines the specific set of instructions provided by the user to display the 3D model more intuitively to the user. Multimodal behavioral features are reintegrated as multimodal data inputs by synthesizing different modal interaction data. They are analyzed by inputting them into the instruction set system and LLM. In the process, the system extracts the corresponding cue words and generates the final answer based on these cue words. Next, based on the user's needs and environmental conditions, the corresponding speech or environment synthesis is per-formed to ensure that the output is adapted to the current interaction

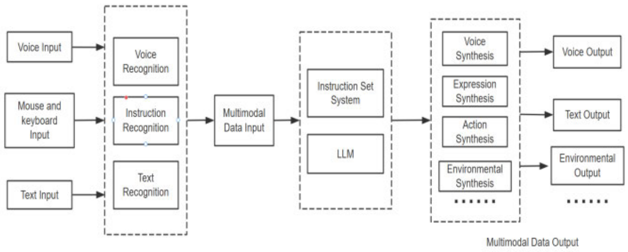


FIGURE 8. Multimodal interaction design.

FIGURE 9. Panoramic view of the west annex of Baoguang Hall.

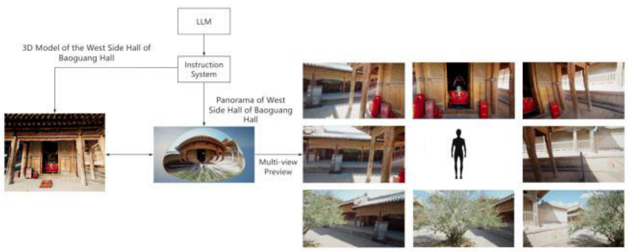


FIGURE 10. Panorama tour interaction.

environment. This content is output to the user interface for the user to intuitively access the desired information or experience.

The specific multimodal interaction design is shown below:

In this paper, we take the west annex hall of Baoguang Hall as an example to synthesize the collected panoramic data and map it to the inner wall of the fisheye sphere to realize the mapping from 3D space to real space. By giving the corresponding commands to the big language model, we can carry out the scene switching and jumping operations in the form of command jumping. The collected panoramic data takes Baoguang Hall as an example, as shown in Figure 9 below, and the entire panoramic mapping and jump process is shown in Figure 10.

In this paper, common natural weather phenomena in Qinghai Province are simulated by changing the geometry and properties of particles, including sunny, cloudy, rainy, snowy, floating dust and foggy. By analyzing the differences in the weather patterns of these six, they are divided into three main

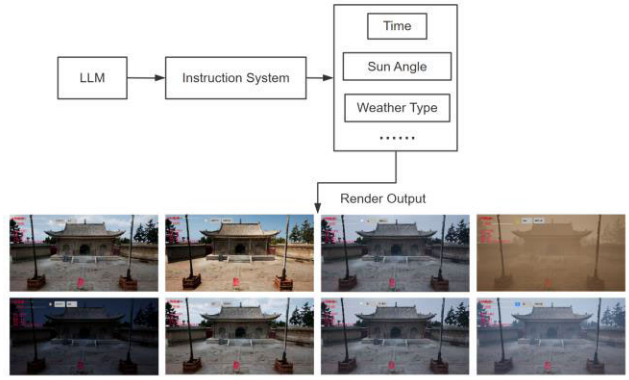


FIGURE 11. Ambient weather interaction.

TABLE 1. Text question-answering performance.

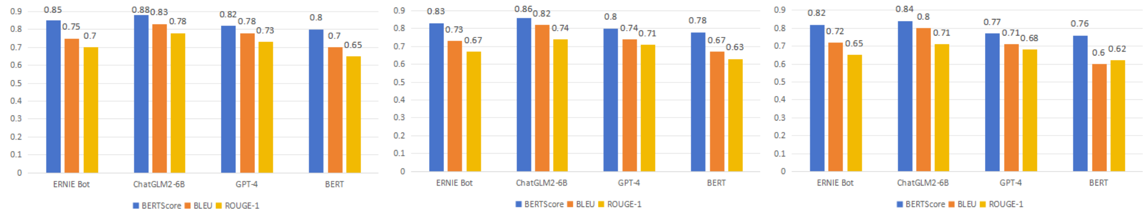
model	Cultural Question-answering Accuracy	BERTScore	Response delay (ms)
TIGMI	95%	0.97	120
ERNIE bot	83%	0.85	125
ChatGPT-4	79%	0.82	150
BERT	75%	0.80	150

categories: sunny and cloudy days without particles, rainy and snowy days with ordinary particles, and floating dust and foggy days with tiny particles. By integrating real-time weather data for environmental simulation, users can freely switch between weather types and view the simulated environment online that matches the local climate. It is also possible to adjust the weather effects and atmosphere of our scenes through commands and related parameters, rendering the weather and time of day that meets the needs of visitors. The weather interactive rendering is shown in Figure 11.

E. EXPERIMENTAL RESULT

We used ERNIE bot, ChatGPT-4 and BERT models as comparison models and analyzed their performance in this scenario. ChatGLM2-6B was fine-tuned to obtain the ChatGLM2-6B-QLoRA-int4 model. It was found that the accuracy of the model was significantly increased compared to the base model. To further validate this finding, we used datasets of 500, 1000, and 2000 sentences along with the three previous metrics to evaluate the model's response quality. Specific experimental results show that on the 500-sentence dataset, the fine-tuned model improves the BERTScore by 0.09, the BLEU by 0.15, and the ROUGE-1 by 0.21. Similarly, the fine-tuned model shows a similar performance improvement on the 1000- and 2000-sentence datasets.

In a comparison of 500 culture-specific questions against the models mentioned above, we found that TIGMI achieved an accuracy of 95% (BERTScore=0.97), which is significantly better than that of the universal LLM. The specific comparison is presented in the following table.



（a）500 Test Questions. （b）1000 Test Questions. （c）2000 Test Questions.

FIGURE 12. Performance on different number of question tests.

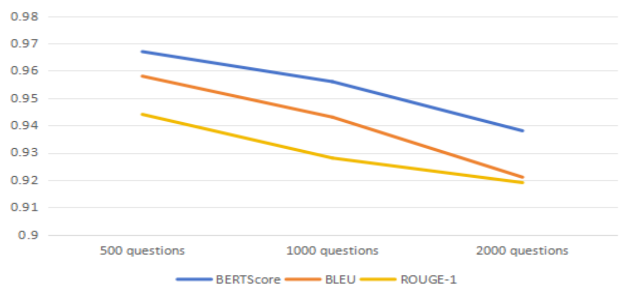


FIGURE 13. Test results after fine-tuning.

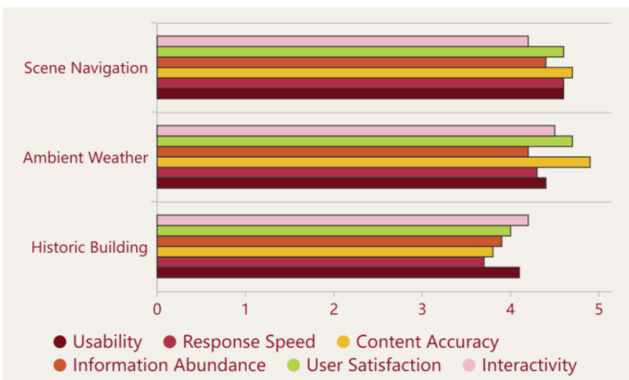


FIGURE 14. User experience of instruction set interaction.

Through user experience evaluation, we comprehensively assessed the system's performance in personalized tour generation. Weighting the five scores of the three aspects of the evaluation, we calculated that all three are above good, and the two types of interactions—environmental weather and scenario guides—barely reach the excellent level, which indicates that the whole system's interactions perform well overall. The monument-building interaction is slightly inferior to the other two interaction methods, but 90% of the users who experienced the project were satisfied and are inclined to recommend the program to others for us.

The fine-tuned model outperformed baseline LLMs across all metrics. By integrating domain-specific databases with real-time RAG, TIGMI ensured 95 percent answer accuracy, nearly 30 percent better than BERT's performance. Leveraging QLORA, the computational cost of fine-tuning LLMs was reduced by 60% while maintaining semantic fidelity, enabling deployment on edge devices. The unified pipeline processes text, audio, and keyboard inputs, generating context-aware outputs with a latency of <150 ms and achieving a user satisfaction score of 4.7/5. However, scene and weather transitions occasionally exceed 200 ms due to network fluctuations, impacting real-time interaction. While user feedback emphasized the value of historical multimodal interactions, architectural interactions require further optimization.

V. DISCUSSION

The experimental results validate the effectiveness of TIGMI in addressing core challenges of personalized meta-cultural tourism, revealing three key findings:

QLoRA-RAG collaborates to improve cultural accuracy. By dynamically injecting domain-specific knowledge through RAG, the framework performs better in generating culturally nuanced questions, such as explaining the symbolism of the Gautama Temple roof design in the absence of a detailed knowledge base for a specific question, which relies on a self-constructed database to give an accurate answer instead of the fuzzy answers given by other large-language models.

Multimodal interaction reduces cognitive load. User satisfaction with the unified text-audio-3D interaction pipeline scored 4.7/5, higher than text-only interfaces, confirming the value of multimodal systems for enhancing virtual tourism engagement. Notably, scene transitions with <150 ms latency meet real-time interaction standards, though weather simulations occasionally exceed 200 ms due to particle system complexity.

TIGMI's modular design can quickly adapt to other cultural sites with limited resources. By simply replacing domain-specific databases, the framework has shown decent narrative capabilities in pilot tests at Huining Temple, showing its potential as a universal tool for digital heritage conservation. By simply replacing domain-specific databases, the framework can quickly adapt to the cultural connotations of other cultural heritage, demonstrating its potential as a universal tool for digital heritage conservation.

While the framework advances cultural interpretation, current models support only Chinese and English. Performance degrades for languages with sparse LLM training data due to embedding mismatches. Future work must address low-resource language adaptation via cross-lingual transfer learning. Additionally, while the LLM pipeline achieves low latency, complex 3D interactions occasionally cause delays. Optimizing particle systems and leveraging edge computing could mitigate this. The framework's performance also depends on domain-specific database quality. Retrieval accuracy drops for sites with fragmented or non-textual records. Integrating multimodal knowledge sources (e.g., audio narratives, 3D annotations) may bridge this gap. In real-world applications, reinforcement learning enables continuous optimization of intelligence-environment interactions, and this optimization capability can also be used to adjust quantization and low-rank decomposition strategies and user interaction paths in QLoRA, but currently TIGMI prioritizes real-time responsiveness in larger scenarios. Therefore, optimizing the rendering performance of large-space 3D scenes is a pre-requisite step for their integration.

For broader applications and future directions, the TIGMI framework may extend beyond cultural tourism, offering value in three emerging areas. As a tool in the field of education, the system can simulate historical events to facilitate immersive classroom experiences by adapting three-dimensional interactive pipelines. As a museum tour tool, integrating TIGMI with AR or VR glasses enables a real-time artifact interpreter that enhances context awareness through RAG. As an enabler of community heritage preservation, a crowdsourcing module allows local communities to contribute oral history or dialect knowledge, enrich domain databases, and promote cultural inclusion. In terms of technical updates and applications, future work will explore lightweight RL-LLM hybrid architectures to test adaptive narrative generation and interaction in delay tolerance scenarios.

VI. CONCLUSION

This study proposes TIGMI, a novel framework integrating quantized Large Language Model (LLM) fine-tuning, retrieval-augmented knowledge injection, and multimodal interaction. The hybrid QLoRA-RAG architecture achieves nearly 95% accuracy in text generation within the Quotidian domain, outperforming generalized large language models by approximately 20%, while reducing GPU memory consumption by 60% through 4-bit quantization and low-rank adaptation. This innovation bridges the gap between computational efficiency and domain-specific knowledge retention. The unified multimodal pipeline supports real-time interaction with latency below 150 ms, enabling seamless integration of text, audio, and 3D scene control. User experience evaluations achieved a satisfaction score of 4.7/5, validating the framework's effectiveness in enhancing historical immersion. Validated at Qutan Temple—a site with limited archival documentation—TIGMI leverages eight segmented 3D models to deliver detailed narratives and immersive interactions, providing a scalable solution for preserving underrepresented cultural heritage.

Future work will focus on further optimizing the framework to ensure efficient real-time inference on mobile devices. Concurrently, the personalized recommendation algorithm will be continuously improved by integrating users' historical behavior, preferences, and social network data to better address personalized needs. We will explore the integration of augmented reality technology with the framework to enable a more immersive virtual travel experience. Beyond cultural tourism, TIGMI's modular design will be extended to applications in virtual education, museum management, and disaster recovery planning. Through continuous evolution of the technological architecture, this study provides a reusable paradigm and practical pathway for cultural heritage preservation in the digital twin era.

ACKNOWLEDGMENT

The authors gratefully acknowledge the supports of the major science and technology projects in Qinghai Province (2024-GX-A3). It also thanks to the Qinghai Provincial Laboratory for Intelligent Computing and Application, the Qinghai Provincial Key Laboratory of Media Integration Technology and Communication and high performance computing center of Qinghai University for providing platform support.

REFERENCES

[1] T. Pencarelli, "The digital revolution in the travel and tourism industry," *Inf. Technol. Tourism*, vol. 22, no. 3, pp. 455–476, Sep. 2020, doi: 10.1007/s40558-019-00160-3.

[2] L. T. De Paolis, S. Chiarello, C. Gatto, S. Liaci, and V. De Luca, "Virtual reality for the enhancement of cultural tangible and intangible heritage: The case study of the castle of corsano," *Digit. Appl. Archaeology Cultural Heritage*, vol. 27, Dec. 2022, Art. no. e00238, doi: 10.1016/j.daach.2022.e00238.

[3] A. Dasgupta, S. Williams, G. Nelson, M. Manuel, S. Dasgupta, and D. Gračanin, "Redefining the digital paradigm for virtual museums," in *Proc. Culture Computing. Interact. Cultural Heritage Arts*, Jan. 2021, pp. 357–373, doi: 10.1007/978-3-030-77411-0_23.

[4] O. B. P. Mah, Y. Yan, J. S. Y. Tan, Y.-X. Tan, G. Q. Y. Tay, D. J. Chiam, Y.-C. Wang, K. Dean, and C.-C. Feng, "Generating a virtual tour for the preservation of the (in)tangible cultural heritage of tampines Chinese temple in Singapore," *J. Cultural Heritage*, vol. 39, pp. 202–211, Sep. 2019, doi: 10.1016/j.culher.2019.04.004.

[5] E. Kanun, A. Alptekin, and M. Yakar, "Documentation of cultural heritage by photogrammetric methods: A case study of Aba's monumental tomb," *Intercontinental Geoinformation Days*, vol. 3, pp. 168–171, Jun. 2021.

[6] K. Hofman, G. Walters, and K. Hughes, "The effectiveness of virtual vs real-life marine tourism experiences in encouraging conservation behaviour," *J. Sustain. Tourism*, vol. 30, no. 4, pp. 742–766, Apr. 2022.

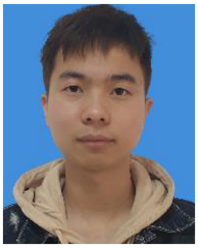
[7] T. Caciora, G. V. Herman, A. Ilies, S. Baias, D. C. Ilies, I. Josan, and N. Hodor, "The use of virtual reality to promote sustainable tourism: A case study of wooden churches historical monuments from Romania," *Remote Sens.*, vol. 13, no. 9, p. 1758, Apr. 2021, doi: 10.3390/rs13091758.

[8] Z. Xie, "The study of Qutan temple in Ledu, Qinghai province," *J. Qinghai Nationalities Inst.*, vol. 1, pp. 20–33, May 1979. [Online]. Available: https://kns.cnki.net/kcms2/article/abstract?v=_GofKS1StuR0y1xyO7tJF gRtjHHhMdsyGlMg0Cep9dgeUuHH2AKbkhSdtC19ahp8MTZ8-tL3oYRB3zqSdx1QqCJK6wmS9ABIn1Y6-7OlowwtjA-2OQWgw2Kv5a3rhFX8VYY9tMfUFT2GLzrV6cQK9F5nyrxbT0ZzTid Boeg2Qt8M8LdH_ldO_6tbGfTcGfOp&uniplatform=NZKPT&language =CHS

[9] A. Baik, "The use of interactive virtual BIM to boost virtual tourism in heritage sites, historic Jeddah," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 9, p. 577, Aug. 2021, doi: 10.3390/ijgi10090577.

[10] H.-M. Chiao, Y.-L. Chen, and W.-H. Huang, "Examining the usability of an online virtual tour-guiding platform for cultural tourism education," *J. Hospitality, Leisure, Sport Tourism Educ.*, vol. 23, pp. 29–38, Nov. 2018, doi: 10.1016/j.jhlste.2018.05.002.

[11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-T. Yih, T. Rocktäschel, S. Riedel, and D. Kiela, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2020, pp. 9459–9474.
[12] Y. Zhou, A. I. Muresanu, Z. Han, and K. Paster. (2023). *Large Language Models Are Human-Level Prompt Engineers*. [Online]. Available: https://openreview.net/forum?id=92gvk82DE-
[13] Z. Jiang, I. Huang, and X. Wang. (2024). *Utilizing Large Language Models for Indoor Tour Guidance*. [Online]. Available: https://deepspatial2024.github.io/papers/DeepSpatial24_paper_6.pdf
[14] R. Alotaibi, A. Ali, H. Alharthi, and R. Almehamdi, "AI chatbot for tourist recommendations: A case study in the city of Jeddah, Saudi Arabia," *Int. J. Interact. Mobile Technol.*, vol. 14, no. 19, pp. 18–30, Nov. 2020, doi: 10.3991/ijim.v14i19.17201.
[15] Z. Wang, L.-P. Yuan, L. Wang, B. Jiang, and W. Zeng, "VirtuWander: Enhancing multi-modal interaction for virtual tour guidance through large language models," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Honolulu, HI, USA, May 2024, doi: 10.1145/3613904.3642235.
[16] Ý. Kýrtýl, B. Çizel, Ý. Uzut, and S. Uzun. (2024). *Bridging the Gap: Fine-Tuning Artificial Intelligence (AI) Chatbots for Tourism*. MTCON, Ýstanbul, Turkey. [Online]. Available: https://avesis.akdeniz.edu.tr/yayin/67a051ca-61a4-4cfc-98a5-e1e5980e766e/bridging-the-gap-fine-tuning-artificial-intelligence-ai-chatbots-for-tourism
[17] T. Mihalic and K. Kuščer, "Can overtourism be managed? Destination management factors affecting residents' irritation and quality of life," *Tourism Rev.*, vol. 77, no. 1, pp. 16–34, Jan. 2022, doi: 10.1108/tr-04-2020-0186.
[18] A. Capocchi, C. Vallone, M. Pierotti, and A. Amaduzzi, "Overtourism: A literature review to assess implications and future perspectives," *Sustainability*, vol. 11, no. 12, p. 3303, Jun. 2019, doi: 10.3390/su11123303.
[19] V. Gowreesunkar and T. Vo-Thanh, "Between overtourism and under-tourism: Impacts, implications, and probable solutions," in *Proc. Overtourism, Causes, Implications Solutions*, H. Séraphin, T. Gladkikh, and T. Vo Thanh, Eds., Cham, Switzerland: Springer, Jan. 2020, pp. 45–68.
[20] S. K. Hoffman, "Online exhibitions during the COVID-19 pandemic," *Museum Worlds*, vol. 8, no. 1, pp. 210–215, Jul. 2020, doi: 10.3167/armw.2020.080115.
[21] F. Besoain, L. Jego, and I. Gallardo, "Developing a virtual museum: Experience from the design and creation process," *Information*, vol. 12, no. 6, p. 244, Jun. 2021, doi: 10.3390/info12060244.
[22] E. C. Ciliberti, M. Fiore, and M. Mongiello, "Development of a metaverse platform for tourism promotion in apulia," in *Proc. IEEE Int. Conf. Metaverse Comput., Netw. Appl. (MetaCom)*, Jun. 2023, pp. 680–681, doi: 10.1109/METACOM57706.2023.00123.
[23] D. Martins, L. Oliveira, and A. C. Amaro, "From co-design to the construction of a metaverse for the promotion of cultural heritage and tourism: The case of Amiais," *Proc. Comput. Sci.*, vol. 204, pp. 261–266, Aug. 2022, doi: 10.1016/j.procs.2022.08.031.
[24] Ş. D. Şirvan and M. Demir, "Exploring digital audio tour guides from visitors' perspective: Benefits and challenges," *Sci. Collection*, vol. 2023, no. 181, pp. 9–17, 2023. [Online]. Available: https://archive.interconf.center/index.php/conference-proceeding/article/view/4857
[25] H. Niu, "The effect of intelligent tour guide system based on attraction positioning and recommendation to improve the experience of tourists visiting scenic spots," *Intell. Syst. Appl.*, vol. 19, Sep. 2023, Art. no. 200263, doi: 10.1016/j.iswa.2023.200263.
[26] I. A. Wong, Q. L. Lian, and D. Sun, "Autonomous travel decision-making: An early glimpse into ChatGPT and generative AI," *J. Hospitality Tourism Manage.*, vol. 56, pp. 253–263, Sep. 2023, doi: 10.1016/j.jhtm.2023.06.022.
[27] R. Hammady, M. Ma, C. Strathern, and M. Mohamad, "Design and development of a spatial mixed reality touring guide to the Egyptian museum," *Multimedia Tools Appl.*, vol. 79, nos. 5–6, pp. 3465–3494, Feb. 2020, doi: 10.1007/s11042-019-08026-w.
[28] Z. Du, Y. Qian, X. Liu, M. Ding, J. Qiu, Z. Yang, and J. Tang, "GLM: General language model pretraining with autoregressive blank infilling," 2021, *arXiv:2103.10360*.
[29] A. Zeng, X. Liu, Z. Du, Z. Wang, H. Lai, M. Ding, Z. Yang, Y. Xu, W. Zheng, X. Xia, W. L. Tam, Z. Ma, Y. Xue, J. Zhai, W. Chen, Z. Liu, P. Zhang, Y. Dong, and J. Tang. (2023). *GLM-130B: An Open Bilingual Pre-Trained Model*. [Online]. Available: https://openreview.net/forum?id=-Aw0rrrPUF
[30] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, "QLoRA: Efficient finetuning of quantized LLMs," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2023, pp. 10088–10115.
[31] O. Topsakal and T. Ç. Akıncı, "Creating large language model applications utilizing LangChain: A primer on developing LLM apps fast," in *Proc. Int. Conf. Appl. Eng. Natural Sci.*, Jul. 2023, vol. 1, no. 1, pp. 1050–1056, doi: 10.59287/icaens.1127.



HANG GUO (Graduate Student Member, IEEE) received the bachelor's degree in management from Jingchu University of Technology, China, in 2021. He is currently pursuing the master's degree in engineering with Qinghai University. His main research interests include artificial intelligence, virtual reality interaction, and visualization.

ZHIQIANG LIU received the M.Eng. degree in information engineering from Muroran Institute of Technology, Japan. He is currently an Associate Professor and a Master's Supervisor with the School of Computer Technology and Application, Qinghai University. His main research interest includes virtual reality.

CHUANQIAN TANG received the master's degree from Qinghai University, in 2019. She was appointed as a Lecturer with Qinghai University, in 2022. Her research interests include virtual reality and visualization.

XIAODAN ZHANG received the degree from the University of Chinese Academy of Sciences. She is currently an Associate Professor and a Ph.D. Supervisor. Her main research interests include multi-source data fusion processing, high-performance computing, and green computing applications. She is a member of the CCF Women's Working Committee. She was the 11th batch of backbone teachers in Qinghai Province, recognized as "Kunlun Talents and High Level Education and Teaching Talents" in Qinghai Province.