Academic Education Publishing House

Higher Education and Practice Vol. 3 No. 5, 2026

Exploring the Construction Pathway of a Digital Cultural Heritage Terminology Database Supported by Large Language Models: A Case Study of Shaanxi Province

Yajie Shen

Xi'an Fanyi University, Xi'an, Shaanxi, China

Abstract: With the increasing demand for standardized cultural heritage translation in global communication, the construction of specialized terminology resources has become increasingly important. Taking Shaanxi Province as a case study, this paper explores the construction pathway of a digital cultural heritage terminology database supported by Large Language Models (LLMs). Drawing on multimodal cultural heritage resources, the study proposes a Human-in-the-Loop (HITL) collaborative mechanism featuring "Term Pre-generation-Translator Verification-Dynamic Revision" and develops a three-dimensional "Entity-Attribute-Relation" architecture within the Trados Terminology Management Platform. A bilingual terminology database containing 500 terms is established, covering major categories of Shaanxi's tangible and intangible cultural heritage. The database supports terminology standardization, knowledge organization, and multimodal resource integration, and demonstrates practical value in translation education, translation practice, and cultural tourism international publicity. This paper provides a replicable framework for digital cultural heritage terminology management and contributes to the international communication of regional culture in the era of artificial intelligence.

Keywords: Large Language Models; Digital Cultural Heritage; Terminology Database; Human-in-the-Loop; Multimodal Resources

1. Introduction

Against the backdrop of the ongoing convergence of digital technology and global communication, the international dissemination of cultural heritage places higher demands on the accuracy and consistency of terminology. As one of the major cradles of Chinese civilization, Shaanxi Province boasts rich cultural heritage resources and has developed cultural symbols with distinct regional characteristics, such as the "Terracotta Warriors" and "Qinqiang Opera." However, in the process of international communication, these core cultural terms frequently encounter challenges such as inconsistent translations, semantic deviations, and even a dilution of their cultural significance, which severely hinder the precise transmission of cultural information and its comprehension by international audiences. How to balance terminological standardization with cultural specificity in cross-linguistic communication has become a pressing issue in the field of cultural tourism international publicity.

Consequently, scholarly research has increasingly focused on the role of Large Language Models (LLMs) in translation and terminology processing. On the one hand, existing research has largely focused on improving model performance in specific translation tasks. For instance, Hendy et al. (2023) enhanced the overall consistency of complex texts by introducing document-level translation strategies[1], while Merx et al. (2024) improved translation accuracy in the process of translating from English to Marathi through the strategic selection of parallel corpora and dictionary entries[4]; Vanroy et al. (2024) further noted that prompt design has a significant impact on translation quality, highlighting the critical role of human-machine interaction in the translation process[5]. On the other hand, Chinese scholars are gradually shifting their research focus from "translation tool optimization" to "reconstruction of translation paradigms." For instance, Wang Huashu and Liu Shijie (2024), as well as Yu Jing and Liu Kanglong (2024), have explored the profound impact of LLMs on translation activities from the perspectives of translator subjectivity and research paradigms[6,10]. Meanwhile, in the context of digital and intelligent transformation, the

integration of Computer-Aided Translation (CAT) tools and terminology databases has deepened within specific vertical domains, such as business, traditional Chinese medicine, and law, underscoring the importance of terminology standardization, dynamic updating, and intelligent matching in cross-cultural communication. A review of the existing literature reveals that most studies remain limited to general-purpose text translation or single-modal corpus processing. For the vertical field of cultural heritage-which combines high specialization with profound cultural depth-there is still a lack of systematic and mature research frameworks.

At the same time, terminology research in vertical fields has begun to incorporate cutting-edge technologies such as natural language processing and multimodal approaches, providing rich practical references for this study. For example, Hou Xilong et al. (2019) built a knowledge service platform for intangible cultural heritage based on linked data[2]; Wang Lin's team (2024) established a terminology database for ceramic craftsmanship using machine learning[7]; and Wei Licai (2025) proposed the application prospects of multimodal LLMs in terminology system construction[8]. In terms of multimodal and human-machine collaborative exploration, Yang Kun et al. (2026) constructed a multimodal terminology database integrating text, images, and video links in the field of new energy vehicles, effectively enhancing the visualization of technical concepts and translation accuracy in cross-cultural collaboration[9]; Ji Linlin (2026), meanwhile, innovatively implemented an "editor-led, human-machine collaboration" mechanism in the construction of a medical terminology database, successfully addressing complex semantic conflicts that cannot be easily handled by automated tools through the deep involvement of domain experts[3]. While the above studies offer significant insights regarding technological implementation and collaborative mechanisms, they often focus on single technical pathways or specific vertical fields. Consequently, an integrated research framework that balances terminology standardization, translation collaboration, and outreach applications has yet to be established in the field of digital cultural heritage. In light of this, this paper introduces LLMs to construct a Human-in-the-Loop (HITL) mechanism centered on "Term Pre-generation-Translator Verification-Dynamic Revision." By integrating multimodal corpus resources with a structured organizational approach, this paper explores the construction pathway of a digital cultural heritage terminology database and examines its application value in cultural tourism international publicity, aiming to provide a practical technical reference and replicable paradigm for the international communication of regional culture.

2. Overall Construction Approach and Current Status of the Shaanxi Digital Cultural Heritage Terminology Database

The construction of the Shaanxi Digital Cultural Heritage Terminology Database is based on the integration of a large-scale multimodal cultural heritage corpus, the semantic processing capabilities of LLMs, and the Trados Terminology Management Platform, forming a complete workflow from resource integration to application output.

At the resource level, the research team systematically collects multimodal materials, including archaeological reports from Shaanxi, local gazetteers, oral history records of intangible cultural heritage, artifact images, and 3D scanning data. This data spans multiple categories, such as historical sites, traditional craftsmanship, and folk culture, thereby providing an authoritative and diverse corpus foundation for the construction of the terminology database. At the technical level, LLMs, with their semantic understanding and cross-modal alignment capabilities, undertake core tasks such as automatic term extraction, candidate translation generation, and semantic relation mining, serving as the technological link connecting data processing, translation collaboration, and knowledge organization. At the platform level, the Trados Terminology Management Platform provides stable tool support for the structured storage of terms, multi-dimensional field configuration, and expert collaborative review, enabling the standardized entry and dynamic management of terminology resources.

Building on these foundations, this paper organizes its discussion around four key stages in the construction of the terminology database. First, terminology resource integration empowered by LLMs, encompassing multimodal corpus conversion and term extraction. Second,

the collaborative translation mechanism for terminology, establishing the HITL paradigm centered on "Term Pre-generation-Translator Verification-Dynamic Revision". Third, terminology quality control and dynamic updating, constructing multi-dimensional evaluation indicators and a continuous optimization process. Fourth, the current status of the terminology database on the Trados platform, presenting the entry scale, classification system, and preliminary application outcomes. These four parts unfold progressively, systematically presenting the full picture of the Shaanxi Digital Cultural Heritage Terminology Database, from underlying corpus processing to upper-level application services. Figure 1 illustrates the construction workflow of the Shaanxi Digital Cultural Heritage Terminology Database.

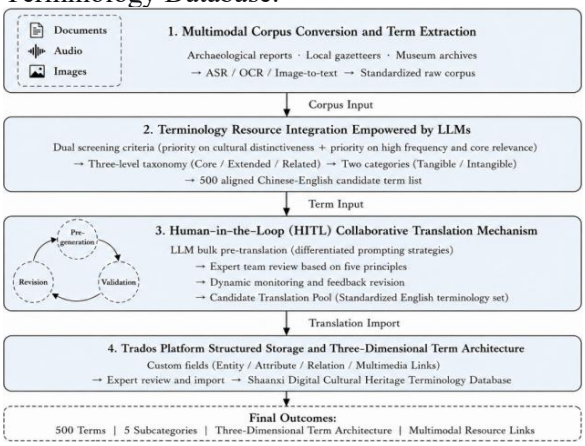

Figure 1. Construction Workflow of the Shaanxi Digital Cultural Heritage Terminology Database

2.1 Terminology Resource Integration Empowered by LLMs

The integration of terminology resources in the field of cultural heritage faces two specific challenges. First, relevant terms are scattered across diverse texts such as archaeological reports, local gazetteers, academic papers, and museum archives. There is a lack of a ready-made centralized term list, making the manual collection and verification process extremely labor-intensive. Second, a substantial amount of cultural information is embedded in non-textual media such as artifact images, intangible cultural heritage performance videos, and dialect audio recordings, making it difficult for traditional term extraction methods to incorporate them into a unified corpus processing workflow. To address these challenges, this paper adopts a data processing approach empowered by LLMs.

In terms of multimodal corpus collection, the primary sources for this study are drawn from the following three categories. The first comprises archaeological excavation reports and local gazetteers, which possess a high degree of academic rigor and historical contextual integrity, serving as the core basis for term definition and provenance. The second comprises museum collection archives and 3D scanning data of cultural relics, which cover the physical attributes, spatiotemporal information, and visual records of precious museum artifacts, providing verifiable empirical material for the attribute dimension of terms. The third comprises intangible cultural heritage performance videos, oral history audio recordings, and dialect audio records. These resources directly capture the living practices of traditional craftsmanship, opera and Quyi, and folk culture, serving as indispensable contextual support for the semantics of terms. After manual screening and expert review, these three categories of resources are organized into a classification catalog based on three dimensions of cultural heritage type, historical period, and geographical region, forming the raw dataset and ensuring the diversity and authority of term sources.

In terms of non-textual data conversion, this paper establishes unified preprocessing standards for heterogeneous data such as images, audio, and video, and executes these standards in parallel. For dialect audio and oral history recordings, Automatic Speech Recognition (ASR) technology is used for transcription to generate text records that can be aligned with the textual corpus. For stone rubbing inscriptions and scanned ancient texts, Optical Character Recognition (OCR) technology is employed to extract textual information. For artifact images and 3D scan data, image-to-text technology is utilized to generate standardized visual descriptions. Building on this foundation, a multimodal data catalog index is developed. Using the JSON-LD format, it records metadata for each non-textual data item-including spatiotemporal attributes, media type, and related terms-establishing an explicit mapping channel between textual and non-textual data. This enables cultural information scattered across different modalities to be aggregated into a unified, structured data stream.

In terms of hierarchical term extraction, this paper designs domain-specific prompt templates tailored to the cultural heritage field to guide LLMs in performing word frequency statistics and domain distribution analysis under context-aware conditions. To ensure the representativeness and usability of the extraction results, this paper establishes dual screening criteria. First, cultural distinctiveness is prioritized, giving preference to terms with strong regional recognition and cultural symbolic value, such as "Mausoleum of the First Qin Emperor," "Qinqiang Opera," and "Fengxiang Clay Sculpture." Second, high-frequency core terms are prioritized; by applying a three-dimensional metric combining weighted word frequency, distribution breadth, and domain authority, high-frequency and stable terms are identified as the core layer. Based on these criteria, the model automatically classifies candidate terms into three levels-core terms, extended terms, and related terms-and categorizes them into the two major categories of tangible cultural heritage and intangible cultural heritage. Taking "Mausoleum of the First Qin Emperor" as an example, this term is classified as a core term, while its associated term "Terracotta Warriors Pits" is categorized as an extended term due to their clear subordinate relationship. In the domain of intangible cultural heritage, "Qinqiang Opera" is classified as a core term, whereas "Banhu" (an accompanying instrument) and "Kuyin" (a vocal branch) are placed in the related term level. This hierarchical classification mechanism transforms terms from isolated entries into structured knowledge units carrying semantic hierarchy and domain affiliation.

Through the systematic integration of the above three levels, this paper ultimately produces a standardized candidate term set covering the major categories of Shaanxi's tangible and intangible cultural heritage, comprising a total of 500 terms. These span five subcategories: ancient sites and tombs, ancient architecture and city walls, precious museum collection artifacts, traditional opera, Quyi and music, and traditional fine arts, handicrafts and folklore. Concurrently, bilingual Chinese-English alignment is completed, providing a clean and complete data foundation for subsequent collaborative translation and structured database entry.

2.2 Collaborative Translation Mechanism for Terminology

Traditional human translation models cannot easily meet the efficiency requirements of large-scale terminology sets, while pure machine translation frequently produces mistranslations regarding culture-specific elements. To address this challenge, this paper constructs a three-stage collaborative translation mechanism-"Term Pre-generation-Translator Verification-Dynamic Revision"-that organically combines the bulk generation capabilities of LLMs with the cultural judgment of domain experts.

In the terminology pre-generation phase, this paper leverages the cross-lingual semantic generation capabilities of LLMs to perform batch pre-translation of the 500 Chinese terms that have been classified and categorized. To address the textual characteristics of cultural heritage terminology, this paper designs a differentiated prompt strategy: for proper nouns such as "Mausoleum of the First Qin Emperor" and "Giant Wild Goose Pagoda," a "transliteration + literal translation + cultural annotation" template is used, requiring the model to simultaneously output Pinyin transcription, literal translation, and a brief cultural explanation; for craft-related terms such as "Qinqiang Opera" and "Fengxiang Clay Sculpture," a "literal translation + functional equivalence + contextual example" template is employed, guiding the model to generate candidate translations alongside usage examples in typical contexts. Once processed by the model, each term yields a Candidate Translation Pool incorporating multiple translation strategies, providing a comparative benchmark for subsequent expert screening.

During the translator verification phase, this study assembles a composite review team comprising translation scholars, linguists, and cultural heritage experts. Based on five translation selection principles-Source-Language Culture Priority, International Convention Priority, Domain Adaptation, Context Differentiation, and Multimodal Correspondence (refer to Section 2.3)-the team reviews each model-generated candidate translation individually. The expert team focuses on evaluating the accuracy of cultural imagery transmission and acceptance in the target language; only terms that pass this review are included in the formal terminology database.

In the dynamic revision phase, this study

establishes a closed-loop update mechanism of "monitoring-feedback-revision." This mechanism triggers terminology revisions through a dual-drive approach of academic literature monitoring and user feedback, transforming the terminology database from a static list of translations into an evolving knowledge system. This update mechanism will be discussed in detail in Section 2.3.

2.3 Principles for Terminology Translation and Dynamic Updates

Ensuring the quality of terminology translation relies on clear principles for selecting translations and a continuous, dynamic maintenance mechanism. This section systematically explains the five principles underlying the translator verification process and describes how dynamic updates are implemented.

The five principles for selecting translations are established in response to recurring issues in the translation of cultural heritage terminology. First, the principle of Source-Language Culture Priority: for terms with strong cultural specificity and no direct equivalents in the target language, phonetic transliteration is retained to preserve cultural distinctiveness. For example, translating Qinqiang as "Shaanxi Opera" conveys only geographical information while losing the historical depth and unique characteristics of the genre embodied in the name Qinqiang; therefore, the primary translation is established as "Qinqiang Opera." Similarly, "Fengxiang Clay Sculpture" is adopted rather than the generic "Shaanxi Folk Sculpture," to preserve its place of origin and craftsmanship characteristics. Second, the principle of International Convention Priority: for terms that have established fixed translations in the international academic community, existing standards are followed to avoid confusion caused by alternative translations. For example, "Terracotta Warriors" is used for the world-renowned Qin dynasty funerary sculptures, and "Tang Sancai (tri-coloured glazed pottery)" for the iconic Tang dynasty ceramics. Third, the principle of Domain Adaptation: architectural terms are based on national standards such as the Dictionary of Chinese Ancient Architectural Terminology and international architectural terminology systems to ensure accuracy in professional contexts. Fourth, the principle of Context Differentiation: while precise, equivalent translations are used for the same term in academic literature, explanatory text may be appropriately added in cultural tourism international publicity materials to balance accuracy and readability. Fifth, the principle of Multimodal Correspondence: ensure consistency between the translated names in the terminology database and associated multimedia texts, such as image captions and video subtitles, to avoid translation conflicts for the same cultural object across different media.

Regarding dynamic updates, this paper establishes a dual-drive triggering mechanism. The academically driven component uses LLMs to periodically scan UNESCO databases, publications by the International Council of Museums (ICOM), and authoritative academic journals. When new high-frequency translations or alternative terms for a given term are detected, it automatically generates a report with update recommendations. The application-driven component collects user feedback on the terminology database from translation teaching in higher education and cultural tourism international publicity practices, incorporating cases of translation disputes into the revision list. Revision recommendations from both channels must undergo review and confirmation by an expert team before implementation. Each update is accompanied by a record of the revision date, basis, and reviewer, ensuring full traceability of the terminology's evolution.

2.4 Current Status of the Terminology Database

Through the progressive implementation of the aforementioned processes-including resource integration, collaborative translation, and quality control-the Shaanxi Digital Cultural Heritage Terminology Database is preliminarily established on the Trados platform. Currently, the terminology database contains a total of 500 terms. It is organized into two major categories-tangible cultural heritage and intangible cultural heritage-and further divided into five subcategories: ancient sites and tombs, ancient architecture and city walls, precious museum collection artifacts, traditional opera, Quyi and music, and traditional fine arts, handicrafts and folklore. Each term is assigned a unique identifier and aligned bilingually in Chinese and English. In terms of entry structure, the terminology database moves beyond the traditional flat list format. Through custom fields,

it achieves the structured storage of three-dimensional information-Entity, Attribute, and Relation-as well as multimodal resource links, providing foundational data support for subsequent knowledge mining and translation applications.

3. The Three-Dimensional Architecture and Multimodal Implementation of the Cultural Heritage Terminology Database

This chapter focuses on the internal structure of the terminology database, detailing how terms are organized, stored, and presented. Traditional terminology databases typically provide only a flat "term-translation" correspondence, making it difficult to accommodate the spatiotemporal information, cultural connotations, and semantic relations inherent in cultural heritage concepts. To address this limitation, this paper utilizes custom fields within the Trados platform to construct a three-dimensional "Entity-Attribute-Relation" architecture and incorporates multimodal resource links, transforming each term into a knowledge unit that integrates textual descriptions, semantic networks, and multimedia contexts.

3.1 Design of the "Entity-Attribute-Relation" Three-Dimensional Architecture

Cultural heritage terminology often involves multiple dimensions of information: a single term refers to a specific cultural object or form of craftsmanship, carries specific spatiotemporal coordinates and cultural value, and is semantically linked to other concepts. A simple "Chinese-English" parallel field cannot meet the requirements for storing and retrieving the aforementioned information. To address this, this paper designs a three-dimensional field structure within the Trados terminology database.

The Entity dimension serves as the core anchor for terminology entries, storing the Chinese name and its English equivalent. The Attribute dimension comprises three subfields: the Temporal-Spatial Feature subfield records dates and geographic information; the Physical Attribute subfield records characteristics such as materials and form; and the Cultural Value subfield records elements such as intangible cultural heritage level and historical significance. The Relation dimension links terms to other relevant terms through semantic association types, forming a traceable knowledge network.

Taking "Mausoleum of the First Qin Emperor" as an example, the entity field stores its Chinese name and the English equivalent "Mausoleum of the First Qin Emperor"; the Temporal-Spatial Feature subfield records "Lintong District, Xi'an / Qin Dynasty"; the Physical Attribute subfield records "the mausoleum complex covers nearly 56 square kilometers"; and the Cultural Value subfield records "the tomb of China's first emperor"; the Relation field points to "Terracotta Warriors Pits," indicating a subordinate relationship. This structure transforms "Mausoleum of the First Qin Emperor" from an isolated entry into a knowledge node carrying multidimensional information.

3.2 Linking Terminology to Multimodal Resources

Accurate understanding of cultural heritage terminology often requires visual and auditory information-it is difficult to fully convey the vocal style of Qinqiang Opera or the imposing grandeur of the Terracotta Warriors through textual descriptions alone. To address this, this paper introduces a "Multimedia Links" field within the three-dimensional architecture, directly linking multimodal resources-such as images, audio, video, and 3D models-to terminology entries.

Multimodal linking primarily employs two methods. First, direct linking, which binds resource files to terms, such as linking a 3D scan model to "Terracotta Warriors Pits" and aerial images of the mausoleum complex to "Mausoleum of the First Qin Emperor". Second, contextual association, which links multiple complementary resources to a single term. For example, "Qinqiang Opera" is linked to audio recordings of Kuyin arias, images of opera masks, and performance videos of classic plays, thereby reconstructing the cultural context of the term from both auditory and visual dimensions.

The implementation of these associations relies on the JSON-LD metadata index described above-each non-textual data item is annotated with spatiotemporal attributes and related terms during the preprocessing stage. Upon import, custom connectors map the metadata to Trados fields, ensuring accurate correspondence between multimodal resources and terms.

3.3 Application Example of the Three-Dimensional Architecture: The Case of

"Qinqiang Opera"

To further illustrate how the three-dimensional architecture is organized in practice, this section uses "Qinqiang Opera" as an example to present a complete, structured entry for this term in the terminology database.

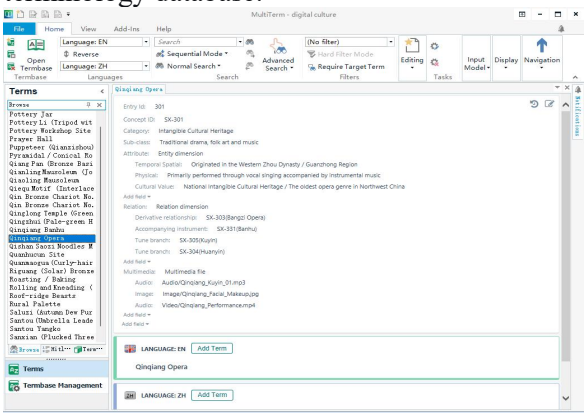

Figure 2. Schematic Diagram of the Three-Dimensional Term Architecture and Multimodal Association Taking "Qinqiang Opera" as an Example

As shown in Figure 2, in the Entity dimension, the Chinese name and its English equivalent "Qinqiang Opera" are recorded. In the Attribute dimension, the Temporal-Spatial Feature records "Originated in the Western Zhou Dynasty / Guanzhong Region"; the Physical Attribute, given that Qinqiang Opera is an intangible cultural heritage item, records "primarily performed through vocal singing accompanied by instrumental music"; and the Cultural Value records "National Intangible Cultural Heritage / The oldest opera genre in Northwest China." The Relation dimension links to "Bangzi Opera" (derivative relationship), "Banhu" (accompanying instrument), "Kuyin" and "Huanyin" (vocal branch), forming a sub-network of opera knowledge centered on "Qinqiang Opera." The Multimedia Link field incorporates audio clips of Kuyin arias, images of Qinqiang Opera facial makeup, and performance videos of classic plays, creating a complementary relationship between the term's textual information and its multimodal context.

This structure ensures that "Qinqiang Opera" is not merely a translation entry in the terminology database, but a knowledge unit that integrates textual descriptions, semantic relations, and multimedia contexts. When working on texts related to Qinqiang Opera, translators can consult this entry to simultaneously access multiple layers of information-including the standardized translation, cultural connotations, vocal style, and performance forms-thereby more accurately grasping the semantic layers of the term and making well-informed translation decisions.

4. The Application Value of the Terminology Database

The value of a terminology database is ultimately realized through its practical application. The resource integration, collaborative translation, and structured database entry described in previous chapters have laid the foundation for the database to be effectively utilized in three key contexts: translation education, translation practice, and cultural tourism international publicity.

In translation education, the terminology database developed in this study has been integrated as a teaching resource into the English Translation Practice course at Xi'an Fanyi University, and students enrolled in the course were granted full access to it. During the course's specialized translation training on cultural heritage, students can directly search for relevant terms within the Trados platform upon receiving translation assignments and use standardized translations from the database, thereby avoiding mistranslations caused by unreliable sources found through online searches. At the same time, students can access the multimodal resources linked to each terminology entry-such as listening to audio recordings of Kuyin arias from Qinqiang Opera to perceive its vocal style, viewing images of the "Guahu" (clay tiger wall hanging) of Fengxiang Clay Sculpture to grasp its visual characteristics, and examining 3D-scanned models of the Terracotta Warriors Pits to understand their spatial layout-thereby compensating for the limitations on translation quality caused by a lack of cultural background knowledge. For example, when translating an introductory text on Qinqiang Opera, students listened to the audio excerpt and, after understanding the auditory characteristics of "Houqiang," no longer simply translated it as "shouting style." Instead, by referencing the standardized translation "Houqiang (roaring vocal style)" in the terminology database, they accurately conveyed the cultural connotations of this unique vocal style. A semester-long pilot program demonstrated that, with the aid of the terminology database, students significantly improved the standardization of translations for

culture-specific terms and deepened their understanding of the cultural connotations behind them. This teaching case has been developed into a scalable course module, providing practical teaching resources and a replicable paradigm for cultivating "translation + technology" interdisciplinary talent in the age of intelligent transformation.

In translation practice, the research team compiled the Bilingual Terminology Handbook of Shaanxi Cultural Heritage based on the terms in the database. The handbook organizes Chinese-English parallel terms with brief annotations into five categories: ancient sites and tombs, ancient architecture and city walls, precious museum collection artifacts, traditional opera, Quyi and music, and traditional fine arts, handicrafts and folklore. The handbook serves as a reference for publicity staff at cultural heritage institutions and translation professionals, resolving the previous issue of inconsistent translations for the same cultural terms across different translation projects. Additionally, the three-stage collaborative mechanism proposed in this paper-"Term Pre-generation-Translator Verification-Dynamic Revision"-can be applied to professional terminology translation tasks of similar scale, providing an operational workflow for HITL translation practice.

In cultural tourism international publicity, Shaanxi, as a major province for cultural tourism, hosts a large number of international visitors annually at cultural heritage sites such as the Terracotta Warriors, the Giant Wild Goose Pagoda, and Huaqing Palace. Consequently, the demand for terminology translation in scenic area commentary, cultural and creative product descriptions, and international cultural events continues to grow. Zhong Weihe (2014), drawing on the perspective of the sociology of translation, explored the construction of a research framework for Culture-Oriented Translation, noting that cross-border cultural exchange and translation activities are, at their core, processes of external communication between different cultural fields mediated by cultural symbols[11]. Inspired by this theory, the digital cultural heritage terminology database developed in this paper serves not merely as a tool for lexical conversion, but as a foundational support for cross-cultural field communication. This terminology database provides a unified reference for translation in the aforementioned practical cultural tourism contexts, reducing cultural information deviations caused by inconsistent translations. Consequently, the terminology database helps overcome reception barriers between source and target cultural fields, thereby facilitating the accurate dissemination and effective construction of Shaanxi's cultural symbols in the international field. As the database expands and its dynamic update mechanism continues to operate, the depth of its application in cultural tourism international publicity is expected to be further enhanced.

5. Conclusion

From the perspective of LLM-empowered terminology management, this paper systematically explores the construction pathway of a digital cultural heritage terminology database for Shaanxi Province. The study constructs a development model characterized by "LLM-driven, HITL collaboration, three-dimensional architecture, and dynamic updating." It proposes a three-stage collaborative translation mechanism of "Term Pre-generation-Translator Verification-Dynamic Revision," formulates five principles for selecting translations-including Source-Language Culture Priority and International Convention Priority-and implements a three-dimensional "Entity-Attribute-Relation" term architecture along with multimodal resource linking within the Trados platform. Through this approach, this study completes a terminology database covering five subcategories of Shaanxi's tangible and intangible cultural heritage, comprising a total of 500 bilingual term pairs, thereby providing standardized terminology support for cultural heritage translation practice and education.

Looking ahead, this research will focus on deepening the application of the terminology database in cultural tourism international publicity through the following two directions. At the application strategy level, first, the terminology database will be integrated with bilingual audio guide systems at scenic spots, providing standardized terminology support for tour commentary and exhibit labels at cultural heritage sites such as the Terracotta Warriors and the Giant Wild Goose Pagoda, ensuring consistent presentation of translations in on-site tour scenarios. Second, the terminology database will be embedded into the translation workflow for cultural and creative products, offering

terminology review services to support the development of such products by cultural and museum institutions, thereby ensuring the accurate transmission of cultural symbols in commercial contexts. Third, drawing on the bilingual resources of the terminology database, terminology training modules will be developed for international-facing cultural tourism professionals to enhance tour guides' and publicity staff's understanding and application of translation standards for culture-specific terms. At the level of international communication significance, the terminology database serves as a foundational linguistic infrastructure for cross-cultural communication. Its core value lies in reducing semantic deviations caused by translation inconsistencies through the standardization of terminology and its presentation in multimodal contexts, thereby ensuring the accuracy of cultural information during cross-linguistic transmission and enhancing the recognition and acceptance of Shaanxi's cultural symbols among target-language audiences. As the terminology database becomes more deeply integrated with cultural tourism international publicity scenarios, it is expected to evolve from a translation aid into a terminology infrastructure for the international communication of regional culture, providing a reference model for the construction of cultural heritage terminology databases and international communication practices in other Chinese provinces and cities.

Acknowledgements

This paper is a interim result of the 2025 Scientific Research Program Funded by Education Department of Shaanxi Provincial Government (Program No. 25JK0120), titled "Research on the Construction of the Shaanxi Digital Cultural Heritage Terminology Database Empowered by Large Language Models and Its Application in Cultural Tourism International Publicity".

This paper was supported by 2025 University-level Scientific Research Project of Xi'an Fanyi University titled "Research on the Construction of the Shaanxi Digital Cultural Heritage Terminology Database Driven by Large Language Models" (Project No. 2025B48).

References

[1] Hendy A, Abdelrehim M, Sharaf A, et al. How good are GPT models at machine translation? a comprehensive evaluation[J]. arXiv preprint arXiv:2302.09210, 2023.

[2] Hou Xilong, Tan Guoxin, Zhuang Wenjie, et al. Research on knowledge management of intangible cultural heritage based on linked data [J]. Journal of Library Science in China, 2019, 45(02): 88–108.

[3] Ji Linlin. Construction practice and application analysis of an applied medical terminology database: A case study of the terminology database in the PPH knowledge digital service system [J]. Publishing Reference, 2026, (1): 16–20.

[4] Merx R, Mahmudi A, Langford K, et al. Low-resource machine translation through retrieval-augmented llm prompting: A study on the mambai language[J]. arXiv preprint arXiv:2404.04809, 2024.

[5] Vanroy B, Lefer M A, Macken L, et al. Proceedings of the 1st Workshop on Creative-text Translation and Technology[C]//Proceedings of the 1st Workshop on Creative-text Translation and Technology. European Association for Machine Translation, 2024.

[6] Wang Huashu, Liu Shijie. The impact of large language models on translator subjectivity and mitigation strategies [J]. Foreign Languages and Translation, 2024, 31(04): 13–19+98.

[7] Wang Lin, Wang Hao, Li Xiaomin, et al. Construction and application of a terminology database for intangible cultural heritage ceramic craftsmanship integrating learning extension [J]. Library Tribune, 2024, 44(02): 66–78.

[8] Wei Licai. Narrative, identity, immersion: Strategies for promoting cultural heritage protection and inheritance empowered by multimodal large models in the new era [J]. Journal of Yunnan Minzu University (Philosophy and Social Sciences Edition), 2025, 42(01): 31–39.

[9] Yang Kun, Wang Mengqi. Construction and application of a multimodal terminology database in the field of new energy vehicles [J]. China Terminology, 2026, 28(1): 126–131.

[10] Yu Jing, Liu Kanglong. Reshaping translation studies: Paradigm shift and future directions under the influence of AI technology [J]. Journal of Foreign Languages (Journal of Shanghai International Studies University), 2024,

47(04): 72–81.

[11] Zhong Weihe, Feng Man. Constructing a research framework for culture-oriented translation from the perspective of the sociology of translation [J]. Foreign Languages Research, 2014, (3): 57–62.