Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

NED-GraphRAG for Evidence-Constrained Traditional Dress Design: Integrating User Needs with Cultural Knowledge Graphs

Chang Liu
Donghua University
Ping Cang
cangping492@gmail.com

Donghua University

Article

Keywords: cultural heritage, traditional dress, Song dynasty, knowledge graph, retrieval-augmented generation, evidence traceability, user-centered design, cultural sustainability

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11014752/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Type of the Paper (Article)

NED-GraphRAG for Evidence-Constrained Traditional Dress Design: Integrating User Needs with Cultural Knowledge Graphs

Chang Liu [1], Ping Cang[1,2]*

[1] College of Fashion and Design, Donghua University, Shanghai, China; 1239110@mail.dhu.edu.cn.
[2] Shanghai Urban Construction Vocational College, Shanghai, China; cangping492@gmail.com
* Correspondence: cangping492@gmail.com

Abstract

Traditional dress design based on historical culture requires evidence that can be used within a specific design context and traced back to its source. Focusing on literati gatherings in the Song dynasty, this study develops NED-GraphRAG, a GraphRAG framework that incorporates contemporary user needs into the retrieval and use of cultural evidence. User needs were identified from 3,052 public online comments using BERTopic, coding assisted by a large language model, and researcher review. The Analytic Hierarchy Process was then conducted with 18 target users to determine the relative importance of cultural elements under three design need orientations: Historical Verification, Visual Communication, and Social Ritual. Evidence drawn from historical texts, images, cultural relic materials, and modern scholarship was standardized into 2,053 evidence units and organized into a cultural evidence graph containing 4,233 nodes, 12,749 relationships, and 854 manually verified semantic triples. Six retrieval methods were evaluated using 30 test queries. NED-GraphRAG achieved an nDCG@10 of 0.821 and an AHP-AS@5 of 0.852, exceeding GraphRAG Retrieval by 0.031 and 0.104, respectively. Claim support precision reached 0.897, and the user task success rate was 92.2%. The results show that user needs can be reflected in evidence priority while the underlying cultural evidence remains unchanged. Graph relations and source controls further support the traceability of evidence used in design recommendations. The framework provides a practical method for using cultural evidence in the contemporary design of traditional dress.

Keywords: cultural heritage; traditional dress; Song dynasty; knowledge graph; retrieval-augmented generation; evidence traceability; user-centered design; cultural sustainability

1. Introduction

Traditional dress is a type of cultural heritage that needs to be understood in its historical context [1]. Its sustainable transmission cannot rely only on the storage of isolated images or descriptive texts. Historical evidence also needs to remain understandable and reusable in contemporary contexts such as education, exhibition, design, and public participation [2]. Generative artificial intelligence has expanded the pathways for cultural heritage preservation and communication. Image generation, virtual reconstruction, and interactive systems can lower the technical threshold for public participation and broaden the circulation of traditional cultural resources [3, 4]. Yet technological intervention does not necessarily bring sustainability to traditional dress culture [5, 6]. When a generation lacks constraints from cultural relationships and recognition of contemporary needs, tra-

ditional dress can be reduced to homogeneous stylistic symbols. This may limit the preservation and sustainable transformation of this form of cultural heritage. A sustainable digital practice therefore needs to build stable connections among reliable evidence, contemporary user needs, and cultural reuse.

Knowledge graph (KG) approaches and retrieval-augmented generation (RAG) provide new technical pathways for this aim. Cultural heritage studies often use ontology frameworks such as CIDOC CRM to describe cultural objects, people, events, and their relationships. These studies also combine entity recognition, relation extraction, and large language model-assisted methods to extract knowledge from historical texts, image descriptions, and cultural relic materials. RAG improves the accuracy of generated content by retrieving external evidence. GraphRAG further organizes entities and their associations through graph structures, supporting cross-source retrieval, multi-hop relation expansion, and relationship interpretation. These methods have been applied in cultural heritage knowledge management, museum display, educational communication, and intelligent question-answering, forming a technical basis for heterogeneous data integration and knowledge use [7, 8]. In the field of traditional dress, related studies can standardize the organization of dress knowledge and support retrieval and knowledge-based question-answering. Work oriented toward design transformation remains limited. The key issue in design generation lies in evaluating the fit among historical evidence, user needs, and use contexts. Traditional dress design generation therefore needs to go beyond data retrieval and further address evidence ranking, relationship interpretation, and claim constraints.

Literati gatherings in the Song dynasty provide a representative setting for examining this issue. As shown in Figure 1, a literati gathering comprises people, events, objects, settings, and meanings and can be understood as a highly relational cultural practice. Its distinctive features also make it an important reference for contemporary culture-tourism integration, museum activities, and traditional cultural experiences [9, 10]. This study therefore selects literati gatherings in the Song dynasty as its research object. It examines whether relational evidence from this context can form an explainable fit with user needs in the contemporary design transformation of traditional dress. This forms the central issue through which the study addresses sustainable transformation.

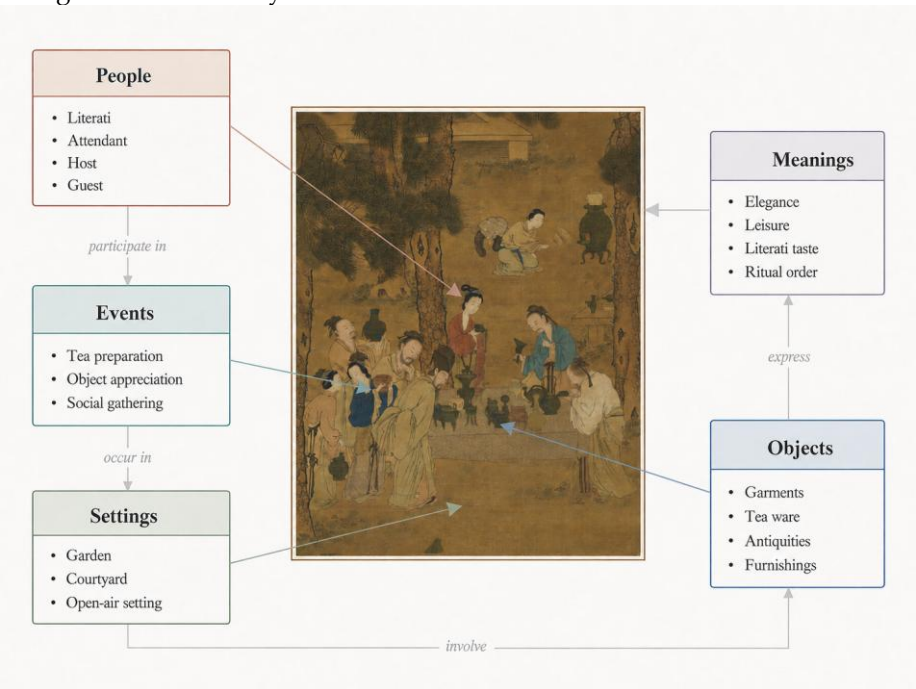


Figure 1. Relational structure of Song dynasty literati gatherings.

Based on the above discussion, this study proposes the following research questions. RQ1: How can dispersed evidence from images, texts, and cultural relics related to traditional dress in literati gatherings in the Song dynasty be organized into a cultural relationship structure that supports design judgment? RQ2: In different design transformation contexts, how can contemporary user perceptions and preferences for cultural elements be incorporated into evidence screening and the generation of design recommendations? RQ3: Can incorporating user need weights, historical evidence, graph relationships, and source constraints into the GraphRAG process improve evidence fit and explainability in traditional dress design generation?

To address these questions, this study proposes an NED-GraphRAG (Need–Evidence–Design Graph Retrieval-Augmented Generation) framework for organizing cultural evidence from Song-dynasty literati gatherings and supporting the design transformation of traditional dress. The study first uses social media comments as the user corpus and applies BERTopic to identify contemporary users' discussion topics around traditional dress and literati gatherings in the Song dynasty. It then constructs LLM-assisted coding prompts based on field theory. The researchers review, revise, and merge the coding results, interpret user concerns through the five cultural dimensions of people, events, objects, settings, meanings, and distill them into three design need orientations: Historical Verification, Visual Communication, and Social Ritual. On this basis, AHP is used to calibrate preferences for cultural elements under different design need orientations. Images of literati gatherings in the Song dynasty, classical texts, cultural relic reports, and modern studies are standardized as traceable evidence units and connected as a five-dimensional cultural evidence graph. In the retrieval stage, NED-GraphRAG retrieves and reranks candidate evidence, then embeds evidence, relationship chains, and usage constraints into design recommendations. This allows dress forms, colors and patterns, display modes, and design prompts to be traced back to specific evidence.

The main contributions of this study are fourfold. First, it connects social-media-based user perception analysis with cultural heritage evidence retrieval. Second, it builds a cultural evidence graph of literati gatherings in the Song dynasty that connects the five cultural dimensions of people, events, objects, settings, and meanings. Third, it proposes an AHP-guided evidence graph reranking method. Fourth, it embeds historical evidence, cultural relationships, and usage constraints into the NED-GraphRAG generation process, forming an evidence-based design decision-making framework for the sustainable transformation of traditional dress culture.

2. Literature Review

2.1. Artificial Intelligence and the Sustainable Transformation of Cultural Heritage

Cultural heritage sustainability responds to the protection mandate set out in SDG 11.4. Its central concern is to sustain physical forms, cultural knowledge, and meanings amid social change [11]. Within this framework, AI contributes by strengthening basic conservation capacities, including heritage resource identification, digital restoration, damage detection, and risk monitoring[12, 13]. Computer vision, for example, can identify material deterioration and structural damage from images and field data. These capabilities help preserve the physical basis needed for cultural resources to enter public communication and contemporary applications[14, 15].

As its applications have extended into public communication, generative artificial intelligence (genAI) has expanded approaches to cultural heritage education and design[16]. Recent studies on genAI in education have shifted from academic integrity and curriculum integration to prompt-design skills. Critical thinking underpins learners' ability to formulate questions, assess outputs, and collaborate with AI [17, 18]. In design

practice, AIGC has primarily changed how tasks are allocated and decisions are made [19]. Studies of New Year paintings and Zisha pottery show that the value of the technology depends on whether the generated design proposals respond to user needs while remaining grounded in reliable cultural evidence [20, 21].

Changes in human-AI relations can heighten the risks associated with the sustainable transformation of cultural heritage. Research has identified potential risks of cultural authenticity loss, homogenization, and copyright disputes arising from AI involvement [22, 23]. However, existing studies generally assess technical effectiveness through user acceptance, willingness to participate, and experience quality. They pay less attention to cultural accuracy, evidence sources, and contextual fit [24, 25]. Recent studies have begun to incorporate user cognition and cultural requirements into generative design[26]. Stable methods for integrating historical evidence and cultural relationships into design judgment remain underdeveloped. The sustainable transformation of cultural heritage therefore requires stronger knowledge organization and evidence traceability.

2.2. Knowledge Organization and Evidence Traceability in Cultural Heritage

Knowledge graphs provide a methodological basis for integrating heterogeneous cultural evidence and interpreting relationships[27]. CIDOC CRM, for example, offers a shared descriptive framework for cultural objects, people, events, time, and places. Previous studies have applied ontologies and knowledge graphs to digital heritage management, relationship retrieval, and knowledge interpretation in traditional Chinese culture [28-30]. Yet knowledge graphs based primarily on textual entities often struggle to represent the full structure of cultural knowledge. Multimodal knowledge graphs have therefore become an important research direction [31, 32]. CICHMKG integrates the categories, texts, and images of intangible cultural heritage projects into a single graph to address the fragmentation of Chinese intangible cultural heritage data [33]. WuMKG establishes a domain ontology for Chinese painting and calligraphy. It combines textual entities with visual features to describe artworks and their artistic relationships [34].

The integration of knowledge graphs with large language models has extended the use of cultural heritage knowledge to question-answering and generation tasks[35]. RAG retrieves external materials through three stages: indexing, retrieval, and generation. This process allows the model to produce answers based on the retrieved content [36]. Text chunking, however, can disrupt relationships across documents. Vector similarity may also retrieve evidence that is semantically close but contextually mismatched, thereby limiting the integration of knowledge from multiple sources [37]. Knowledge graph-enhanced RAG uses entity relationships to organize text chunks. Retrieval can then expand along relationship paths, while the graph structure provides information for evidence reranking. Document GraphRAG further combines a document relationship graph with a domain entity graph to improve cross-paragraph retrieval in complex documents. Its evaluation has focused primarily on question answering in the manufacturing domain [38].

In the cultural heritage domain, systems for Nanjing Yunjin brocade, traditional folk customs, and ancient dress have applied knowledge graphs to RAG and intelligent question answering [39-41]. These applications still focus primarily on factual queries and knowledge presentation. Design transformation requires retrieval results to be filtered according to evidence sources, cultural relationships, and use contexts. This requirement has yet to be adequately addressed.

2.3. Relational Perspectives on Contextual Adaptation in Heritage Design

Field theory offers a relational approach to understanding cultural heritage design. Bourdieu situates cultural practices within the interaction among field, habitus, and capital. He argues that the meanings of cultural works depend on the conditions of their production, circulation, and reception [42, 43]. In sustainability research, this theory is often used from a relational perspective to examine the conditions and underlying logic of transformation. Relevant studies argue that sustainable transitions are unlikely to be achieved through a single technological improvement or a change in individual attitudes. Technological interventions need to be translated into sustained practices within specific contexts [44–46]. Cultural heritage research further shows that cultural meanings are continually produced through curation, visits, and organizational operations. Contexts become sites in which cultural practices are reorganized [47, 48]. From this perspective, contextual adaptation in heritage design depends on connecting cultural relationships with specific needs and developing sustainable experiences of use [49].

User needs are a key point of contact through which cultural relationships enter design decisions. Related studies often view users as active participants and argue that digital narratives, cultural spaces, and information practices should respond to their cognitive, emotional, and behavioral needs[50]. Tourism research further shows that contemporary experiences of traditional dress are closely associated with cultural motivations and destination settings[51, 52]. In heritage design, the Kano model, AHP, TOPSIS, and Kansei engineering have been used to elicit user needs, determine indicator weights, and evaluate cultural and creative product proposals [53–55]. In research on Chinese cultural heritage digitization and design, this relational perspective is often operationalized through multidimensional analytical frameworks organized around people, events, objects, settings, and meanings. These frameworks describe the relationships among participants, activities, material elements, contexts, and cultural meanings [56, 57].

These studies offer a methodological basis for need-oriented design. Yet most focus on proposal evaluation, product optimization, or knowledge presentation. The fit between design evidence and specific contexts remains underexamined. This study therefore incorporates user needs, cultural relationships, and graph-based evidence into a single screening process. This provides a methodological basis for the contextualized design transformation of traditional dress at Song-dynasty literati gatherings.

3. Research Methodology

The study comprises four stages, as illustrated in Figure 2. In the first stage, online comments related to Song-dynasty literati gatherings, dress, and aesthetics are collected. BERTopic is used for semantic topic modeling. Field theory, LLM-assisted interpretation, and researcher review are then combined to derive the dimensions of user perception and need orientations. Based on these results, AHP calibrates evidence preference weights for the three need orientations, providing a basis for subsequent evidence ranking. The second stage constructs a cultural knowledge graph of Song-dynasty literati gatherings from historical paintings, classical texts, and reports on cultural relics. The graph organizes dispersed materials as queryable and verifiable relational evidence. In the third stage, NED-GraphRAG retrieves and reranks candidate evidence and organizes the resulting evidence context. This process allows evidence retrieval to account jointly for query content, cultural relationships, source reliability, and need orientation. The selected evidence is then organized into structured design recommendations covering dress forms, colors and patterns, presentation modes, and usage constraints. The fourth stage evaluates the framework through retrieval evaluation, graph audit, generation evaluation, claim-level evidence audit, ablation experiments, and task-based user evaluation. All methods involving

human participants were carried out in accordance with relevant guidelines and regulations. Informed consent was obtained from all individuals who participated in the AHP survey, expert evaluation, and user task evaluation.

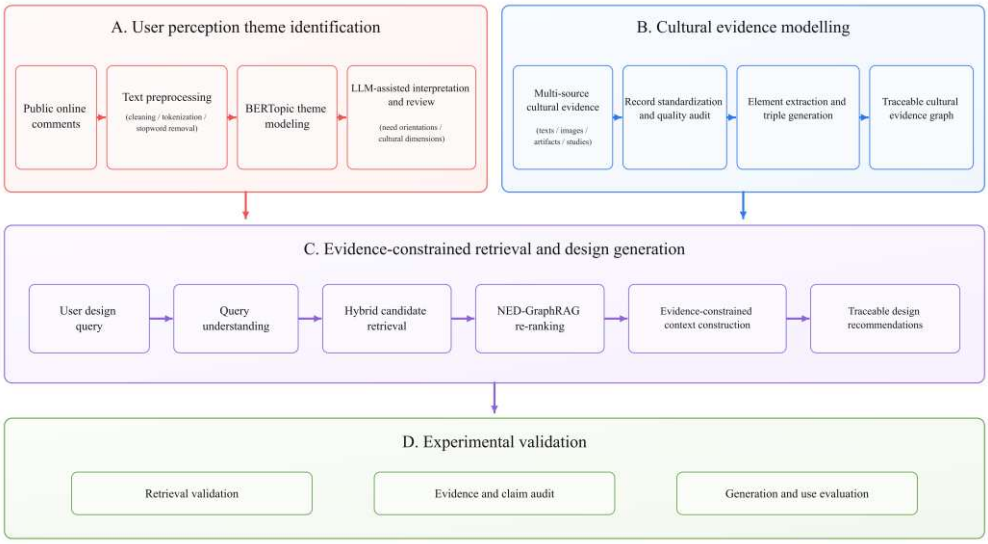


Figure 2. Research workflow of the NED-GraphRAG framework.

3.1. Contextual User Need Modeling

3.1.1.Identification of User Perceptions

User needs in traditional culture design research are commonly derived from researcher experience, interviews, or questionnaires. These approaches facilitate the development of interpretive frameworks. Their results may still be affected by limited sample sizes and subjective judgment. Online comments provide a relatively direct account of contemporary audiences' perceptions and aesthetic expectations regarding the dress culture of the Song dynasty. This study therefore used online comments to identify differences in needs across contemporary contexts in which traditional dress is viewed, worn, photographed, and experienced.

Public comments on Song-dynasty literati gatherings, Song-dynasty dress, and Song-style Hanfu were collected from Bilibili, Xiaohongshu, and Douyin using Octoparse. The data covered the period from 2020 to 2026. After irrelevant texts were removed through manual screening, the final corpus contained 3,052 valid comments. Data collection was limited to publicly visible comment text. Usernames, account IDs, IP addresses, and other personally identifiable information were not collected. Preprocessing was conducted in Python 3.14. The main procedures included text deduplication, link removal, deletion of emojis and special characters, removal of @mentions, normalization of repeated characters, and exclusion of meaningless short sentences. Chinese word segmentation was then performed using Jieba. A custom domain lexicon was developed using terms related to the dress of the Song dynasty, literati gathering activities, image titles, and the language used in online comments. This lexicon reduced the risk of incorrectly segmenting domain-specific expressions such as "宋代雅集" (literati gatherings in the Song dynasty), "褙子" (beizi), and "团扇" (round fan). Stopword removal was based on general Chinese stopword resources, including the Harbin Institute of Technology stopword list. The list was manually revised using high-frequency, low-information terms identified in the corpus. Function words, modal particles, platform-specific expressions, and online expressions with little semantic content were removed.

During topic modeling, BERTopic was applied to the cleaned comment corpus for semantic clustering. LDA relies primarily on word-frequency co-occurrence, whereas

BERTopic uses a pretrained language model to generate semantic embeddings. This makes BERTopic more suitable for online comments characterized by short sentences, colloquial language, and varied formulations. Let the cleaned comment corpus be $C = \{c_1, c_2, \ldots, c_N\}$. Each comment is mapped to a vector representation by a pretrained semantic model:

$$\mathbf{z}_i = f_\phi(\mathbf{c}_i), \quad \mathbf{z}_i \in \mathbb{R}^d \tag{1}$$

where c_i denotes the ith comment, f_ϕ denotes the pretrained semantic encoding model, and z_i is the semantic vector of the comment. UMAP was then used to reduce the dimensionality of the vector space, and HDBSCAN was used to cluster semantically similar comments [58, 59]. For each resulting topic, BERTopic used c-TF-IDF to extract representative keywords. The weight of each keyword is expressed as:

$$w_{t,k} = tf_{t,k} \times \log\left(1 + \frac{A}{df_t}\right) \tag{2}$$

where $w_{t,k}$ denotes the weight of term t in topic k, $\mathrm{tf}_{t,k}$ denotes the normalized frequency of term t in topic k, df_t denotes the number of topics containing the term, and Adenotes the average term count across topics. MMR was subsequently applied to rerank candidate keywords and reduce redundancy.

After the initial BERTopic topics were obtained, the representative keywords and typical comments for each topic were entered into an LLM to assist with the generation of topic labels and semantic interpretations. The researchers reviewed these results against the original comments. The topics were then classified using the five-dimensional cultural framework of people, events, objects, settings, and meanings. These classifications were further consolidated into three need orientations: Historical Verification, Visual Communication, and Social Ritual. Table 1 presents the BERTopic topics and their interpretive classifications.

Table 1. BERTopic themes, cultural dimensions, and need orientations.

Topic ID	BERTopic theme	Keywords	Cultural dimensions	Need orientations
T0	Historical authenticity and dynastic distinctions	Southern Song, Northern Song, Neo-Confucianism, Tang dynasty, dynastic comparison	events/people	Historical Verification
T1	Dress forms and class-based dress	Hanfu, skirts, xuanqun, fastening ties, Song-style dress, class-based dress	objects/people	Visual Communication / Historical Verification
T2	Commodity economy and commercial practices	shops, merchants, trade, cost, commercial communication	objects/settings	Social Ritual / Visual Communication
T3	Neighboring ethnic groups and cross-dynastic comparison	Liao and Jin dynasties, Vietnam, Mongols, Huaxia, Ming dynasty	events / meanings	Historical Verification

T4	Film and television characters and literary imagery	talented scholars, screen portrayals, literary imagery, character visuals	meanings / people	Visual Communication
T5	Traditional ritual systems and marital ethics	weddings, Three Cardinal Guides, benevolence and righteousness, ritual performance, female identity	events / people	Social Ritual / Historical Verification
T6	Song dynasty tea experiences and cultural gaps	diancha, tea culture, tea settings, reconstruction experiences	events / settings	Social Ritual

3.1.2.AHP-Based User Preference Weighting

To align subsequent retrieval ranking with different design needs, this study further applied AHP to calibrate preferences for cultural elements across three need contexts[60]. The survey focused on target users, primarily traditional dress consumers, Hanfu enthusiasts, and individuals with experience of immersive cultural tourism. Under scenarios corresponding to Historical Verification, Visual Communication, and Social Ritual, each respondent conducted pairwise comparisons among five cultural elements: people, events, objects, settings, and meanings.

For any need context p, the pairwise comparisons provided by each respondent form the following judgment matrix:

$$\mathbf{A}^{(p)} = \left(a_{ij}^{(p)}\right)_{5\times5} \tag{3}$$

where $a_{ij}^{(p)}$ denotes the importance of the ith cultural element relative to the jth cultural element under need context p, subject to:

$$a_{ji}^{(p)} = \frac{1}{a_{ij}^{(p)}}, \quad a_{ii}^{(p)} = 1 \tag{4}$$

Saaty's 1–9 scale was used to compare the importance of the five cultural elements under each need context. Table 2 presents the scale definitions and response rules.

Table 2. Saaty's 1–9 scale.

ScaleValue	Definition	Explanation
1	Equal importance	The two cultural elements contribute equally to the given demand scenario.
3	Slightly more important	The left-side element is slightly more important than the right-side element.
5	Clearly more important	The left-side element is clearly more important than the right-side element.
7	Strongly more important	The left-side element is strongly more important than the right-side element, with a significant difference in importance.

9	Extremely more important	The left-side element is extremely more important than the right-side element, indicating the highest level of dominance.
2、4、6、8	Intermediate values	These values are used when the judgment falls between two adjacent importance levels.
1/n	Reciprocal values	If the right-side element is more important than the left-side element, the reciprocal value is used, where n = 2, 3, …, 9. A smaller value indicates stronger importance of the right-side element.

Following the pairwise comparisons, each respondent produced a 5×5 judgment matrix for each need context. The geometric mean for the ith cultural element is calculated as:

$$\mathrm{GM}_i = \left(\prod_{j=1}^{n} a_{ij} \right)^{\frac{1}{n}} \tag{5}$$

The normalized weight is then calculated as:

$$w_i = \frac{\mathrm{GM}_i}{\sum_{j=1}^{n} \mathrm{GM}_j} \tag{6}$$

where $n = 5$, and $w_i$denotes the relative weight of the i th cultural element under the current need context. A consistency test was conducted for each judgment matrix. The consistency index is calculated as:

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1} \tag{7}$$

The consistency ratio is calculated as:

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}} \tag{8}$$

The criteria layer contains five cultural element dimensions, giving a judgment matrix of order 5 and an average random consistency index of RI =1.12. A judgment matrix passed the consistency test when $CR \leq 0.10$.

AHP judgments were collected from 18 traditional dress consumers and enthusiasts. Eighteen initial judgment matrices were obtained for each need context. After consistency testing, 16 matrices were retained for each context. The judgments of multiple respondents were aggregated using the geometric mean to form a group judgment matrix for each need context. The group weights of the five cultural elements were then calculated and normalized as:

$$W_p = \{w_{p,\mathrm{people}}, w_{p,\mathrm{event}}, w_{p,\mathrm{object}}, w_{p,\mathrm{setting}}, w_{p,\mathrm{meaning}}\} \tag{9}$$

where:

$$\sum_{k=1}^{5} w_{p,k} = 1 \tag{10}$$

Table 3 presents the resulting AHP weights. These weights are used to calculate W_{need} during NED-GraphRAG evidence reranking.

Table 3. User Demand Profiles and Cultural Element Weights

Need orientation	people	events	objects	settings	meanings	CR
Historical Verification	0.1716	0.1028	0.2819	0.0878	0.3559	0.0016
Visual Communication	0.1894	0.1028	0.3962	0.2494	0.0621	0.0111
Social Ritual	0.1686	0.4038	0.0984	0.2666	0.0625	0.0087

3.2. Evidence Corpus and Five-Dimensional Cultural Knowledge Graph Construction

3.2.1. Cultural Evidence Corpus

The cultural evidence corpus was compiled primarily from classical texts, images of literati gatherings in the Song dynasty, archaeological and cultural relic materials, and modern scholarship. The classical sources included representative texts concerning narratives of literati gatherings, dress institutions, and urban life in the Song dynasty. These texts included Xiyuan Yaji Tu Ji (Record of the West Garden Literati Gathering), the Yufu Zhi (Treatise on Carriages and Dress) in the Song Shi (History of Song), Dongjing Meng Hua Lu (A Dream of Splendor in the Eastern Capital), and Meng Liang Lu (A Record of the Millet Dream). Because Meng Liang Lu is extensive, only volumes concerning textiles, dress materials, and urban consumption were selected. Dongjing Meng Hua Lu was used primarily for its accounts of folk customs, banquets, leisure activities, and urban life.

The image sources included extant paintings of literati gatherings, Wenhui Tu (Literary Gathering), Tingqin Tu (Listening to the Qin), and related figure paintings from the Song dynasty. Long handscrolls such as Emperor Huizong's Eighteen Scholars and Liu Songnian's Nine Elders were divided according to the section numbers used in the National Palace Museum's digital collection. This arrangement preserved details of figures, dress, objects, and settings. Archaeological and cultural relic materials supplied evidence concerning surviving dress objects, tomb finds, materials, and craft techniques. Modern scholarship assisted in interpreting image subjects, dress institutions, and cultural contexts.

Before records were added to the corpus, each item was documented by title, historical period, source type, provenance, and text passage or image identifier. The records were then organized into evidence tiers according to their source characteristics. Scanned PDFs, image-based documents, and selected image descriptions were processed using ABBYY FineReader 16. The OCR output was manually corrected and converted into searchable text. Editable files in formats such as DOCX and TXT were extracted and standardized directly. To prevent filenames, directory information, and low-quality metadata from entering the knowledge graph, a graph_enabled flag was assigned during preprocessing. Only records containing usable body text or reliable descriptive text entered the knowledge extraction and triple generation process. The remaining materials were retained as catalog records. Table 4 reports the corpus scale and audit results.

The corpus contains two evidence layers: graph-linked evidence and the full retrieval corpus. The first layer consists of a manually curated graph-extraction inventory containing 147 source records. These comprise 105 image-related records, 29 historical texts, and 13 reports on cultural relics or dress. Following review, 59 records met the criteria for inclusion in the graph. Content-hash deduplication consolidated 869 raw fragments into

691 graph-linked evidence units. A graph-linked evidence unit is an evidence unit that can be explicitly connected to an existing graph node, visual tag, or source record. It does not denote a manually verified triple. The second layer is the full retrieval corpus. A further 388 usable files were processed. After duplicate files, system-control documents, and materials with low relevance were excluded, 1,362 additional evidence units were added. The final retrieval corpus contains 2,053 evidence units. The curated knowledge graph serves as the relational backbone and includes 854 manually verified semantic triples. At the graph database level, it contains 4,233 nodes and 12,749 links among sources, entities, tags, and evidence.

Table 4. Composition of Evidence Units in the Full Retrieval Corpus.

Evidence category	Evidence units in the full corpus	Graph-linked evidence units	Primary design function
Historical texts	549	549	Dress institutions, activities, rituals, urban life, and cultural meanings
Cultural relic and dress evidence	260	112	Evidence concerning structure, materials, craft techniques, patterns, and objects
Painting and image-description evidence	133	26	Evidence concerning figures, silhouettes, colors, poses, objects, and settings
Modern academic research	1111	4	Interpretive, terminological, and comparative support
Total evidence units	2053	691	An integrated basis for retrieval and relational evidence

Each evidence unit was assigned a content hash and a canonical source identifier to prevent multiple copies of the same file from receiving excessive ranking weight. Different digital copies of the same article or image description were merged at the semantic level. Each long handscroll was modeled as a single artwork record linked to multiple section images and local regions. Descriptions shared by the entire work were indexed once at the artwork level. Visual features and local annotations were retained separately for each section or local region.

3.2.2. Five-Dimensional Cultural Knowledge Graph

After the cultural evidence had been modeled, the usable evidence records were converted into knowledge graph data. Each evidence record is represented as:

$$e_i = (s_i, t_i, y_i, l_i, x_i, m_i, q_i) \tag{11}$$

where s_i denotes the canonical source identifier, t_i the title, y_i the historical period, l_i the page number or image region, x_i the textual or visual content, m_i the metadata set, and q_i the quality status. The set of semantic triples is represented as:

$$T = \{(h, r, t) \mid h, t \in V,\ r \in R\} \quad (12)$$

where h and t denote the head and tail entities, respectively; r denotes the relationship; V is the set of nodes; and R is the set of relationships.

Graph construction combined rule-based extraction with LLM-assisted extraction. Rule-based extraction was used to identify stable metadata, work titles, historical periods, personal identities, and explicit terminology. LLM-assisted extraction was used to identify contextual entities, relationships, visual elements, and information on usage constraints. Each candidate triple retained its evidence identifier and extraction method. Low-confidence results were excluded from retrieval until they had undergone manual review. Table 5 presents the main relationship types and examples of their triple representations.

Table 5. Overview of Cultural Evidence Data

Relationship type	Example triple	Meaning	Role in design translation
Participates in	(Scholar, participates in, listening to the qin)	Links personal identity to an activity at a literati gathering	Provides a basis for character definition and identity positioning through dress
Occurs in	(Listening to the qin, occurs in, waterside pavilion)	Links an activity to its spatial setting	Provides a basis for coordinating settings and assessing the atmosphere conveyed by dress
Uses color	(Image evidence, uses color, moon white / azurite blue)	Records color information found in images or texts	Provides evidence constraints for color schemes
Expresses meaning	(Literati gathering setting, expresses meaning, refined elegance)	Links a setting to its aesthetic meaning	Provides a basis for cultural interpretation and stylistic control
Associated with pattern	(Dress or object, associated with pattern, plum blossom / lotus pattern)	Links patterns to dress, objects, or image evidence	Provides image-based or object-based evidence for pattern selection

After construction, the graph was exported as JSON data and Neo4j import files for subsequent evidence retrieval, relationship expansion, and source tracing. The final graph contains 854 manually verified semantic triples. At the Neo4j database level, it contains 4,233 nodes and 12,749 relationships. In addition to the semantic triples, these relationships include auxiliary links connecting source records, evidence fragments, category assignments, visual tags, and evidence traceability information. The number of database relationships therefore exceeds the number of manually verified semantic triples.

This graph structure provides the basis for relationship expansion and evidence reranking in NED-GraphRAG. It allows queries to locate relevant evidence along relationship paths connecting people, activities, dress, settings, and cultural meanings. A representative subgraph of the resulting cultural knowledge graph is shown in Figure 3.

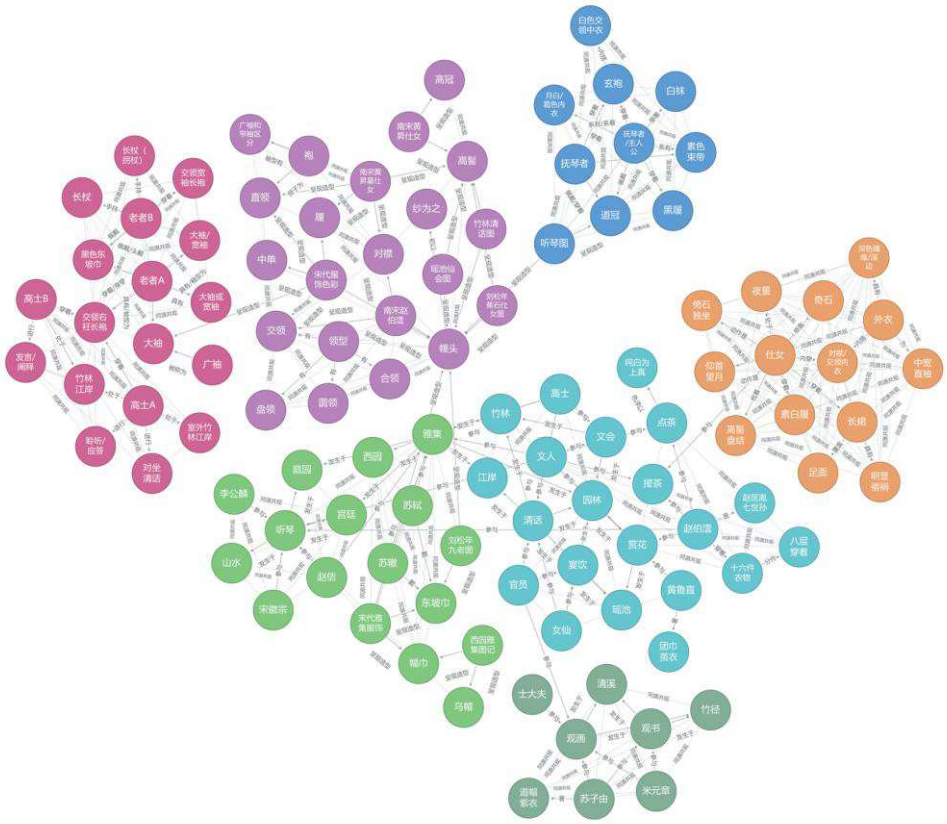


Figure 3. Visualization of the cultural knowledge graph.

3.3. NED-GraphRAG Framework

A set of dress design recommendations may require visual evidence from paintings, contextual evidence about activities from historical texts, material evidence from cultural relic reports, and interpretive support from modern scholarship. Combining multiple sources is therefore necessary in practice. The NED-GraphRAG framework connects user queries, need orientations, the five-dimensional cultural knowledge graph, and evidence-constrained generation. For a specific design context, the system selects combinations of cultural evidence that can support dress design judgments. Query representation, hybrid retrieval, need matching, graph-relation reasoning, and source reliability are then used to organize the selected evidence into a structured evidence context for generation.

3.3.1. Query Representation and Candidate Retrieval

A user query q is represented by a need orientation p, a five-dimensional need vector $u(q)$, and a set of query anchors A_q. The need orientation p corresponds to one of three design contexts: Historical Verification, Visual Communication, or Social Ritual. The vector $u(q)$ describes the intensity of the query's needs across the five cultural dimensions of people, events, objects, settings, and meanings. The anchor set $A_q$extracts retrievable cues such as personal identity, activity, dress requirements, spatial setting, aesthetic meaning, gender, season, and intended use. In the experimental tasks, the need orientation was specified in advance by the task scenario. In the prototype system, it was selected by the user.

Candidate evidence was retrieved through a hybrid strategy combining BM25, Chinese semantic embeddings generated with BAAI/bge-base-zh-v1.5, character n-gram matching, and a visual-metadata channel. BM25 matched historical terms, dress names, and activity names. Dense semantic retrieval handled expressions with similar meanings,

while character matching reduced the effects of segmentation differences and linguistic variation. The visual channel used image descriptions, image sources, and manually annotated labels for dress forms, colors, patterns, and materials. It was therefore limited to metadata-assisted image-text retrieval and did not involve pixel-level encoding of visual regions. The scores from all retrieval components were normalized before fusion to produce the candidate evidence set D_q. The Chinese semantic encoding process used in the dense retrieval channel is illustrated in Figure 4.

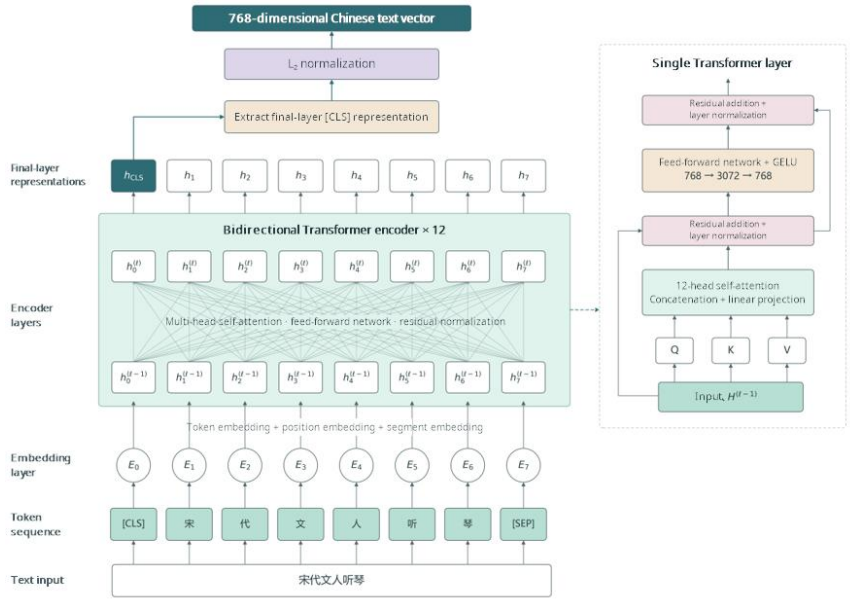


Figure 4. Structure of the BGE Chinese text encoder.

3.3.2. Reranking Based on Need Matching, Graph Relations, and Source Reliability

Following candidate retrieval, NED-GraphRAG reranks evidence according to need matching, graph relations, and source reliability. The need-matching score measures whether an evidence unit corresponds to the priorities assigned to cultural dimensions under a particular need orientation:

$$S_{\text{need}}(q,d,p) = \sum_{k=1}^{5} w_{p,k}\, u_k(q) r_k(d) \tag{13}$$

where $w_{p,k}$ is the AHP weight of the kth cultural dimension under need orientation p, $u_k(q)$ is the intensity of query q for that dimension, and $r_k(d)$ is the degree to which evidence unit d covers the dimension. This formulation allows the three need orientations to affect evidence ranking while retaining the constraints of the specific query. The source content of the evidence corpus remains unchanged across need orientations. The differences occur in evidence priority and the organization of the evidence context.

The graph-relation score measures relationship reachability and path quality between candidate evidence and the query anchors. Starting from A_q, the system expands candidate evidence through one-hop and two-hop relationships. The graph-relation score is defined as:

$$S_{\text{graph}}(q,d) = \max_{\pi \in \Pi(A_q,d)} \eta^{|\pi|-1} \prod_{e \in \pi} o_e \tag{14}$$

where $\Pi(A_q, d)$ denotes the set of admissible paths from the query anchors to evidence unit d, π denotes a relationship path, π denotes its length, $o_e$is the weight assigned to

relationship e, and η is the path-decay coefficient. Direct semantic relationships among people, activities, dress, settings, and meanings receive higher weights, while general co-occurrence relationships receive lower weights. Paths that violate temporal, identity, or source constraints are excluded from the admissible path set.

Source reliability was encoded as an operational scoring rule for evidence ranking, as shown in Table 6. Two researchers independently assigned reliability scores according to source type, historical period, institutional provenance, locatability, and evidential directness. When their judgments differed, the lower score was retained, and the disagreement was recorded. Direct records and modern interpretations from the same source were separated into different evidence units. This prevented interpretive text from inheriting the reliability score of the original material. Unverified online descriptions and materials of unknown origin were excluded during evidence ingestion. They did not enter candidate retrieval, reranking, or claim–evidence binding.

Table 6. Operational Source Reliability Scores Used in Evidence Reranking

Evidence type	Typical sources	Score	Primary evidential role
Archaeological and artifact evidence	Excavated objects, securely documented cultural relics, and material evidence	1.00	Direct support for materials, structure, craft techniques, and object attributes
Historical image evidence	Securely dated paintings and institutional image records	0.95	Visual support for dress forms, colors, patterns, poses, and settings
Historical textual evidence	Primary historical texts and institutional records	0.90	Contextual support for dress institutions, activities, identities, rituals, and usage conditions
Modern academic research	Peer-reviewed scholarly studies	0.72	Interpretive support for terminology, historical interpretation, and cross-source comparison

After the scores from all channels had been calculated, they were normalized and combined into a final reranking score. The score incorporates lexical relevance, dense semantic similarity, character similarity, visual metadata, graph relations, need matching, and source reliability:

$$S(q,d,p) = \lambda_1 S_{\mathrm{BM25}} + \lambda_2 S_{\mathrm{BGE}} + \lambda_3 S_{\mathrm{char}} + \lambda_4 S_{\mathrm{visual}} + \lambda_5 S_{\mathrm{graph}} + \lambda_6 S_{\mathrm{need}} + \lambda_7 S_{\mathrm{source}} \quad (15)$$

Here, S_{BM25}denotes lexical relevance; S_{BGE} denotes dense semantic similarity; S_{char}denotes the character n-gram matching score; S_{visual} denotes the match between the query and image descriptions or visual labels; S_{graph} denotes the graph-relation score; S_{need} denotes the need-matching score; and S_{source} denotes the source reliability score.

In the current implementation, the coefficients for BM25, BGE similarity, character similarity, visual metadata, the graph-relation score, AHP-based need matching, and source reliability were 0.23, 0.27, 0.06, 0.06, 0.15, 0.13, and 0.10, respectively. The coefficients were determined through a stratified grid search using an independent set of 10

development queries. The 30 test queries were excluded from parameter tuning. The development objective

$$J = 0.40\,nDCG@10 + 0.25\,P@5 + 0.20\,SourceCoverage - 0.15\,DuplicateRate \quad (16)$$

Raw coefficient values were searched in increments of 0.05. Each candidate coefficient vector was normalized before evaluation so that the channel weights summed to 1. The search required the source reliability weight to be at least 0.05 and limited the weight of any single channel to 0.30. These constraints reduced the influence of any single relevance measure while retaining source diversity and evidence verifiability. To examine parameter dependence, the subsequent sensitivity analysis varied the core weights assigned to BGE, graph relations, AHP-based need matching, and source reliability above and below their selected values.

3.3.3. Diversity Control and Evidence Context Construction

The reranked candidate evidence is not passed directly into the generation context in its entirety. The same article, image description, or category of modern interpretation may produce several highly similar fragments. The system therefore applies source-aware maximal marginal relevance, referred to as source-aware MMR, to select the $Top-k$ evidence units. This process balances query relevance against evidence diversity. The jth selected evidence unit is defined as:

$$d_j = *\,arg\,max_{d \in D_q \setminus D_{\text{selected}}} \left[\mu S(q,d,p) - (1-\mu) \max_{d' \in D_{\text{selected}}} \text{Sim}_{\text{source}}(d,d') \right] \quad (17)$$

where D_{selected} denotes the set of evidence units already selected for the evidence context, $\text{Sim}_{\text{source}}(d,d')$ denotes the combined source and textual similarity between two evidence units, and μ controls the balance between relevance and diversity. This mechanism prevents near-duplicate descriptions and materials from the same source from occupying multiple positions in the evidence context.

A set of traditional dress design recommendations generally requires evidence from paintings, historical texts, cultural relic reports, and modern scholarship. Paintings provide evidence concerning poses, colors, and settings. Historical texts establish activity contexts and identity constraints. Cultural relic reports provide evidence concerning materials, craft techniques, and structure, while modern scholarship supports terminology, interpretation, and cross-source comparison. Combining evidence from multiple source types is therefore necessary for this design task.

The selected evidence units are organized into a compact evidence context. Each entry contains a source identifier, evidence fragment, image region, relationship path, visual elements, source reliability, and usage constraints. Evidence is grouped by design field. This structure allows the generation process to distinguish among tasks concerning dress forms, colors, patterns and materials, presentation modes, and usage constraints. It also preserves evidence traceability between each design claim and its source. Figure 5 summarizes the shared query–evidence encoding and AHP-guided retrieval fusion process used in NED-GraphRAG.

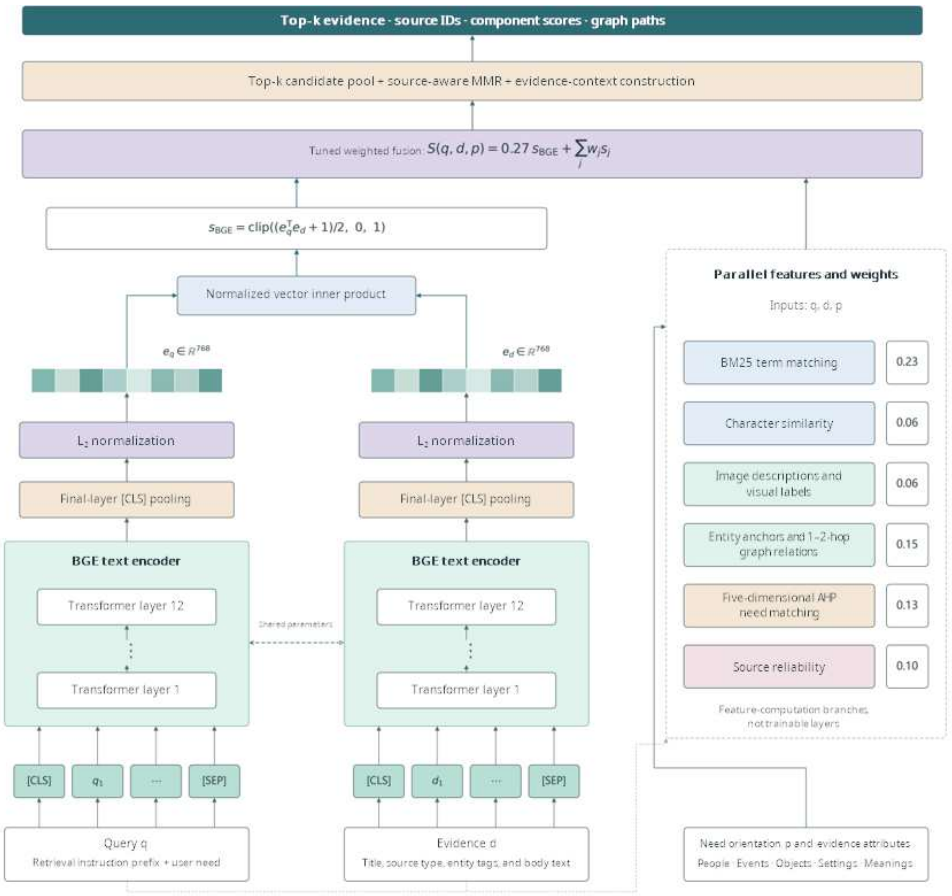


Figure 5. NED-GraphRAG retrieval fusion structure.

3.4. *Experimental Design*

To evaluate NED-GraphRAG in terms of evidence retrieval, generation quality, evidential support, and practical use, this study conducted retrieval comparisons, expert evaluation, claim-level evidence audits, knowledge graph quality audits, ablation and sensitivity analyses, and task-based user evaluation.

3.4.1. Test Queries and Expert Gold Standard

The test set contained 30 representative queries, with 10 queries for each need orientation, covering different design purposes. The Top-20 results returned by the six retrieval methods were pooled and deduplicated, producing 694 query–evidence unit candidate pairs. Three expert reviewers with backgrounds in traditional dress history, fashion design, and digital cultural heritage independently evaluated the candidates. They were blinded to the retrieval method that produced each candidate. Each candidate was rated on a scale from 0 to 2 across three dimensions: query-content relevance, historical appropriateness, and need-orientation fit. For query-content relevance, scores of 0, 1, and 2 indicated irrelevant evidence, a partial response, and a direct response to the core requirements, respectively. For historical appropriateness, the three scores indicated that the evidence was unsupported, uncertain, or clearly applicable. For need-orientation fit, they indicated no fit, partial fit, or strong fit with the objectives of Historical Verification, Visual Communication, or Social Ritual. The final rating for each dimension was the median of the three reviewers' scores. When all three scores differed, the reviewers returned to the original sources and determined the final label through discussion. Inter-reviewer agreement was assessed using $Krippendorff's\ \alpha$,and the distribution of samples across the three rating levels was reported.

3.4.2. Comparison Methods and Retrieval Metrics

The retrieval experiment compared six methods, as summarized in Table 7. BM25 RAG, Character RAG, Dense RAG, and Hybrid RAG represented lexical retrieval, character-enhanced retrieval, dense semantic retrieval, and hybrid retrieval, respectively. GraphRAG Retrieval extended Hybrid RAG by incorporating entity relationships and one-hop and two-hop graph paths. NED-GraphRAG further introduced AHP-based need matching, image metadata, source reliability, source-aware MMR, and structured evidence organization. Plain LLM did not return ranked evidence and was included only in the expert evaluation of generated design proposals.

Table 7. Configuration of the Comparative Retrieval Methods.

Method	Lexical retrieval	Semantic or character matching	Graph relations	AHP-based need matching	Source and diversity control
BM25 RAG	Yes	No	No	No	No
Character RAG	Yes	Character n-gram	No	No	No
Dense RAG	No	BGE	No	No	No
Hybrid RAG	Yes	BGE	No	No	No
GraphRAG Retrieval	Yes	BGE	Yes	No	Partial
NED-GraphRAG	Yes	BGE	Yes	Yes	Yes

To prevent the AHP need orientations used by NED-GraphRAG from being incorporated into the relevance standard, P@5, pooled Recall@5, MRR, and nDCG@10 were calculated using expert-adjudicated query-content relevance alone. Let the aggregated content-relevance rating be $r_{q,i} \in \{0,1,2\}$. The binary relevance label is defined as:

$$y_{q,i} = \begin{cases} 1, & r_{q,i} \geq 1, \\ 0, & r_{q,i} = 0. \end{cases} \tag{18}$$

P@5 is the proportion of relevant evidence among the top five results. Pooled Recall@5 uses all relevant evidence for the same query in the pooled Top-20 candidate set from the six methods as its denominator. MRR is the reciprocal rank of the first relevant result. nDCG@10 retains the graded relevance ratings from 0 to 2 and uses $2^{r_{q,i}} - 1$ as the graded gain. Expert ratings of historical appropriateness and need-orientation fit were retained as separate audit fields and were excluded from the retrieval qrels.

AHP alignment score at 5, denoted as AHP-AS@5, was used to examine the alignment between the ranked results and the current AHP-weighted need profile. Let $s_{q,i}^{\mathrm{AHP}}$ denote the continuous weighted matching score of the ith evidence item. AHP-AS@5 is defined as:

$$\text{AHP-AS@5} = \frac{1}{5}\sum_{i=1}^{5} s_{q,i}^{\mathrm{AHP}}. \tag{19}$$

Because AHP-AS@5 shares the AHP weights used in NED-GraphRAG reranking, it was treated solely as a diagnostic measure of the ranking mechanism. It was not regarded as an independent expert gold standard and was not used alone to support causal claims about method superiority. Independent evaluation of need fit was based on expert ratings of generated design proposals, with method names concealed. Evidence structure was

evaluated separately using source-type coverage, within-result source duplication rate, and valid relationship-path rate.

For each metric, the query-level mean and 95% confidence interval based on 20,000 bootstrap resamples were reported. Differences between NED-GraphRAG and GraphRAG Retrieval were evaluated using a query-level paired bootstrap and a two-sided exact sign-flip randomization test. Holm correction was applied to the three comparisons.

3.4.3. Evaluation of Generated Design Proposals and Evidence Quality

The evaluation of generated design proposals compared four representative methods: Plain LLM, Hybrid RAG, GraphRAG Retrieval, and NED-GraphRAG. Three reviewers evaluated anonymized text proposals generated for the 30 test queries in a randomized order, producing 360 method–query–reviewer evaluation units. A seven-point scale was used to assess historical credibility, query relevance, evidence traceability, need-orientation fit, design applicability, visual coherence, and the clarity of usage constraints. Method names, retrieval scores, and system information were concealed during evaluation. The evaluation was completed in two sessions.

A separate linear mixed-effects model was fitted for each of the seven evaluation criteria:

$$Y \sim Method + Profile + Method \times Profile + (1|Reviewer) + (1|Query) \quad (20)$$

In the model, Method denotes the generation method, Profile denotes the need orientation, and Reviewer and Query represent the random effects of reviewers and queries, respectively. Fixed effects were evaluated using likelihood-ratio tests, and Holm correction was applied to pairwise comparisons. Inter-reviewer reliability was assessed using a two-way random-effects, absolute-agreement, average-measures intraclass correlation coefficient, ICC(2,k).

The evidence audit at the level of individual claims covered the three methods that used retrieval augmentation and returned evidence for their design claims. Sixty design claims were sampled from each method, giving a total of 180 claims. Three reviewers independently assessed each design claim based on five binary indicators covering claim support, evidence coverage, incorrect evidence binding, conflict between claims and evidence, and inference label accuracy. Claim support precision was defined as:

$$CSP = \frac{N_{full} + N_{partial}}{N_{bound}} \quad (21)$$

where N_{full} and N_{partial} denote the numbers of fully supported and partially supported claims, respectively, and N_{bound} denotes the total number of claims linked to evidence. The audit also calculated the rate of claims without evidence, the rate of incorrect evidence binding, the rate of claims that conflicted with the evidence, the accuracy of inference labels, and the evidence coverage of design claims.

The quality of the knowledge graph was audited through stratified sampling. A total of 180 triples were sampled according to relationship type. The sample covered major relationships involving people participating in activities, people wearing dress, activities occurring in settings, dress using colors or patterns, settings expressing meanings, evidence support, and source tracing. Each triple was evaluated for entity correctness, relationship correctness, and source support.

3.4.4. Ablation Analysis, Sensitivity Analysis, and User Task Evaluation

In the ablation analysis, BGE, AHP, graph relations, visual metadata, source reliability, and MMR were removed separately to assess the contribution of each component. The sensitivity analysis examined the four core weights assigned to BGE, graph relations, AHP, and source reliability. Each weight was increased and decreased by 20%, while the

remaining weights were proportionally renormalized. nDCG, source coverage, and duplication rate were reported for each setting.

The user task evaluation recruited 30 participants. Each participant completed tasks corresponding to all three need orientations, with the task order randomized. A scale with seven response points was used to evaluate usability, information clarity, perceived credibility, design usefulness, fit with the selected need orientation, traceability, and intention to continue using the system. Task success required participants to complete a query, switch the need orientation, locate at least one evidence chain, and copy a structured design proposal.

4. Results

4.1. NED-GraphRAG Prototype System and Knowledge Graph Quality Audit

Based on the preceding modeling of user needs, organization of cultural evidence, and graph retrieval method, this study developed an NED-GraphRAG prototype for the design of traditional dress for literati gatherings in the Song dynasty. The system comprises a data layer, a logic layer, and a presentation layer. Figure 6 shows the overall architecture.

The data layer integrates cultural evidence units, passages retrieved through RAG, knowledge graph nodes and relationships, visual elements, and weights associated with each need orientation. These resources form a shared evidence base for retrieval, relationship expansion, and design generation. The logic layer performs query interpretation, candidate evidence retrieval, relationship expansion, reranking based on need orientation, selection of the Top-(k) evidence units, and evidence context construction. It organizes evidence fragments, relationship paths, visual information, and usage constraints as structured input for generation. The presentation layer displays design recommendations and their supporting evidence, allowing users to trace generated results to the original sources.

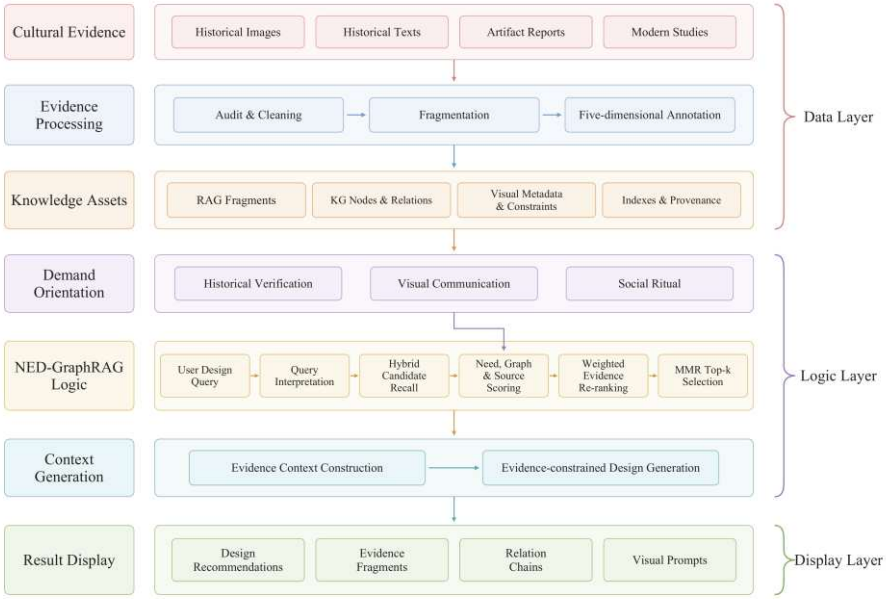


Figure 6. System architecture of the NED-GraphRAG prototype.

The system supports three need orientations: Historical Verification, Visual Communication, and Social Ritual. All three orientations use the same cultural evidence corpus and knowledge graph but apply different weights to the five cultural dimensions during retrieval and evidence context construction. After selecting a need orientation, the system adjusts the ranking priorities assigned to people, events, objects, settings, and meanings.

It also extracts retrieval constraints from the query, including personal identity, activity, dress requirements, setting, and aesthetic meaning. The same query can therefore produce different combinations of evidence and different design priorities while retaining a consistent evidence base.

The prototype uses a workspace interface. Figure 7 presents a representative case involving listening to the qin. The left panel lists the retrieved evidence and its sources. The central panel displays the case image, relationship paths, and visual labels. The right panel presents the structured design proposal. The proposal is organized by dress form, colors and patterns, setting and activity, cultural interpretation, and information on usage constraints. It also retains the correspondence between each design claim and its evidence identifier. Users can inspect the evidence fragments and relationship chains supporting a specific recommendation and verify their sources and applicable contexts.



Figure 7. Representative case interface of the prototype system.

The cultural knowledge graph used by the system contains 4,233 nodes, 12,749 relationships, and 854 semantic triples. It covers people, activities, dress, objects, settings, meanings, and source records. Relationships for evidence support and source tracing are also retained. The quality audit of 180 triples selected through stratified sampling showed entity accuracy of 95.0%, relationship accuracy of 91.1%, source-binding completeness of 93.3%, temporal label accuracy of 92.2%, and identity-constraint accuracy of 88.9%. Agreement among the three reviewers on relationship correctness yielded a $Krippendorff's\ \alpha$ of 0.817. Table 8 presents the audit results by relationship type. Relationships concerning colors used in dress and dress worn by people had relatively lower precision. The main sources of error were fading in paintings, partial occlusion, and differences in the interpretation of dress names across modern studies. These relationships were therefore assigned lower confidence weights during retrieval.

Table 8. Stratified Quality Audit of the Cultural Knowledge Graph.

Relationship type	Sample size	Correct	Precision
Person participates in activity	25	23	0.920
Person wears dress	25	22	0.880
Activity occurs in setting	25	23	0.920
Dress uses color	25	21	0.840
Dress is associated with pattern	20	18	0.900
Setting expresses meaning	20	18	0.900
Evidence support	20	19	0.950
Source tracing	20	20	1.000

4.2. Retrieval Experiment Using the Expert Gold Standard

Among the 694 candidate pairs of queries and evidence units, the numbers assigned relevance scores of 0, 1, and 2 were 191, 231, and 272, respectively. The corresponding distributions were 142, 237, and 315 for historical appropriateness, and 188, 254, and 252 for fit with the selected need orientation. Krippendorff's α was 0.790, 0.824, and 0.756 for the three dimensions, respectively. Agreement was highest for historical appropriateness and relatively lower for fit with the selected need orientation.

Table 9 reports the pooled relevance metrics and the AHP alignment diagnostic. NED-GraphRAG produced the highest point estimates for P@5, pooled Recall@5, MRR, and nDCG@10. Compared with GraphRAG Retrieval, P@5 increased by 0.013 (95% CI [−0.033, 0.060], p = 0.774 after Holm correction), while nDCG@10 increased by 0.031 (95% CI [−0.0039, 0.0654], p = 0.194 after Holm correction). Neither difference was statistically significant. AHP alignment at 5 increased by 0.104 (95% CI [0.0661, 0.1427], p < 0.001 after Holm correction). Because this metric uses the same AHP weights as the reranking process, it serves only to verify whether the ranking mechanism operated in the intended direction.

Table 9. Pooled relevance retrieval performance and AHP alignment diagnostic

Method	P@5	Pooled Recall@5	MRR	nDCG@10	AHP-AS@5 (diagnostic)
BM25 RAG	0.900	0.273	0.933	0.771	0.650
Character RAG	0.900	0.273	0.933	0.772	0.663
Dense RAG	0.913	0.277	0.961	0.731	0.704
Hybrid RAG	0.907	0.275	0.950	0.782	0.721
GraphRAG Retrieval	0.913	0.279	0.950	0.790	0.748
NED-GraphRAG	0.927	0.282	0.978	0.821	0.852

Table 10 further reports the evidence structure metrics for the same test queries and candidate pool. NED-GraphRAG achieved a source type coverage of 0.742, compared with 0.500 for GraphRAG Retrieval. Its source duplication rate decreased from 0.310 to 0.023, while the valid relation path rates were 0.997 and 1.000, respectively. NED-GraphRAG therefore produced broader source coverage and less duplication while maintaining a high proportion of valid relation paths.

Table 10. Evidence structure quality across retrieval methods

Method	Source type coverage	Source duplication rate	Valid relation path rate
BM25 RAG	0.508	0.287	0.620
Character RAG	0.508	0.283	0.634
Dense RAG	0.408	0.370	0.662
Hybrid RAG	0.508	0.293	0.711
GraphRAG Retrieval	0.500	0.310	1.000
NED-GraphRAG	0.742	0.023	0.997

4.3. Expert Evaluation and Evidence Audit

Figure 8 presents the blinded expert ratings for the four methods. NED-GraphRAG received the highest mean score in all seven dimensions, with scores ranging from 5.23 to

5.90. Compared with GraphRAG Retrieval, the largest differences were found in fit with the selected need orientation (Δ = 0.933), evidence traceability (Δ = 0.811), and clarity of usage constraints (Δ = 0.689). Differences in the other four dimensions ranged from 0.389 to 0.567. After Holm correction, significant differences remained for evidence traceability, need-orientation fit, and clarity of usage constraints. Plain LLM received scores of 2.10 for evidence traceability and 2.80 for clarity of usage constraints. Both were lower than its scores for query relevance and visual coherence.

Across Historical Verification, Visual Communication, and Social Ritual, the mean ratings of NED-GraphRAG for fit with the selected need orientation were 5.67, 5.80, and 5.73, respectively. The ICC(2,k) values across the seven dimensions ranged from 0.766 to 0.952. Evidence traceability and clarity of usage constraints had ICC values of 0.952 and 0.942, respectively.



Figure 8. Heatmap of mean expert ratings.

The evidence audit at the level of individual design claims further examined the correspondence between generated content and cited evidence. Plain LLM was excluded because it did not produce explicit bindings between design claims and evidence. Table 11 presents the results for the three methods that incorporated evidence. NED-GraphRAG achieved a claim support precision of 0.897, compared with 0.849 for GraphRAG Retrieval and 0.830 for Hybrid RAG. Its evidence coverage reached 0.967, while the rates of claims without evidence and incorrect evidence binding both decreased to 0.033. No claims directly contradicted their bound evidence. Krippendorff's α values across the five binary audit indicators ranged from 0.589 to 0.846 across the three reviewers. The accuracy of inference labels for NED-GraphRAG was 0.900.

Table 11. Audit of evidence binding for individual design claims across three methods

Method	Claim support precision (supported/bound)	Rate of claims without evidence	Incorrect evidence binding rate	Conflicting claim rate	Evidence coverage
Hybrid RAG	0.830(39/47)	0.217	0.133	0.033	0.783
GraphRAG Retrieval	0.849(45/53)	0.117	0.083	0.017	0.883
NED-GraphRAG	0.897(52/58)	0.033	0.033	0.000	0.967

4.4. Ablation and Weight Sensitivity Analysis

The ablation results are presented in Table 12. Removing the graph relation module reduced nDCG@10 from 0.821 to 0.782 and P@5 from 0.927 to 0.893. This was the largest decline in relevance metrics among the ablation settings. Removing the AHP need weights reduced AHP alignment at 5 from 0.852 to 0.741. Removing source reliability reduced source type coverage from 0.742 to 0.675. By comparison, removing BGE similarity or visual metadata produced smaller changes.MMR primarily affected evidence structure. Its removal increased nDCG@10 and P@5 to 0.836 and 0.933, respectively. However, source type coverage decreased from 0.742 to 0.525, while the source duplication rate increased from 0.023 to 0.270. The local improvement in relevance metrics was therefore not accompanied by an improvement in source structure.

The sensitivity analysis varied each of the four weights for BGE, graph relations, AHP, and source reliability by 20%. The remaining weights were proportionally renormalized. Across the perturbed settings, nDCG@10 ranged from 0.833 to 0.852, AHP alignment at 5 ranged from 0.828 to 0.861, source type coverage ranged from 0.708 to 0.758, and the source duplication rate ranged from 0.023 to 0.061. Spearman rank correlation coefficients between the perturbed settings and the baseline ranking ranged from 0.91 to 0.97. NED-GraphRAG maintained its overall ranking position relative to Hybrid RAG under all perturbations.

Table 12. Module ablation results based on the expert gold standard.

Model configuration	nDCG@10	P@5	AHP-AS@5 (diagnostic)	Source type coverage	Source duplication rate
Full NED-GraphRAG	0.821	0.927	0.852	0.742	0.023
Without BGE similarity	0.808	0.913	0.838	0.733	0.027
Without AHP need weights	0.794	0.907	0.741	0.767	0.040
Without graph relation score	0.782	0.893	0.812	0.758	0.027
Without visual metadata	0.814	0.920	0.831	0.742	0.023
Without source reliability	0.807	0.913	0.844	0.675	0.017
Without source-aware MMR	0.836	0.933	0.849	0.525	0.270

4.5. User Task Evaluation

Thirty participants completed 90 tasks across the three need orientations, with an overall task success rate of 92.2%. As shown in Table 13, the mean ratings for all seven measures exceeded 5. Fit with the selected need orientation received the highest overall rating at 5.83, while perceived trustworthiness received the lowest at 5.18.

Among the three task types, Historical Verification received the highest mean for information clarity at 5.93. Visual Communication received the highest mean for fit with the selected need orientation at 5.93, while Social Ritual received the highest mean for design usefulness at 5.77. Evidence traceability was rated similarly across the three task types, with means of 5.57, 5.53, and 5.53. The relatively high ratings for information clarity, design usefulness, and evidence traceability were not accompanied by a comparable rating for perceived trustworthiness.

Table 13. Mean user task ratings across three need orientations.

Need orientation	Ease of use	Information clarity	Perceived trustworthiness	Design usefulness	Fit with need orientation	Evidence traceability	Continued use intention
Historical Verification	5.50	5.93	5.37	5.70	5.87	5.57	5.50
Visual Communication	5.70	5.70	5.07	5.63	5.93	5.53	5.23
Social Ritual	5.60	5.70	5.10	5.77	5.70	5.53	5.43
Overall	5.60	5.78	5.18	5.70	5.83	5.54	5.39

5. Discussion

The full NED-GraphRAG configuration showed the most consistent overall performance in retrieval relevance, fit with need orientation, source diversity, and generation constrained by evidence. Plain LLM produced relatively complete design proposals, but its design claims lacked stable source support. Fluent prose and detailed visual descriptions did not ensure adequate evidential support. This pattern echoes earlier concerns about source opacity, weakened cultural authenticity, and contextual mismatch when AI is applied to cultural heritage [5, 6, 22]. The results for BM25 RAG, Dense RAG, and Hybrid RAG show that external evidence retrieval can strengthen the correspondence between generated content and historical materials. Retrieval therefore provides a foundation for factual grounding and source tracing [36]. The higher valid relation path rate achieved by GraphRAG Retrieval shows that graph structures can establish structured connections among queries, entities, and evidence units. Previous studies on cultural heritage knowledge graphs have likewise organized dispersed texts, images, and entities into searchable relational networks [33, 34]. Graph relations alone did not resolve source diversity or contextual selection. Compared with GraphRAG Retrieval, NED-GraphRAG produced higher results for fit with need orientation, source coverage, and claim support. Cultural heritage design therefore requires additional controls for user needs, source reliability, and evidence diversity. The ablation results clarify the function of each module. Removing AHP reduced AHP alignment at 5 from 0.852 to 0.741, indicating that the need weights primarily affected alignment with the selected need orientation. Removing graph relations reduced nDCG@10 to 0.782, confirming their contribution to evidence ranking. Removing MMR slightly increased nDCG@10 to 0.836. Source type coverage then decreased to 0.525, while the source duplication rate increased to 0.270. A narrow emphasis on relevance can therefore reduce evidence diversity.

The main contribution of this study lies in extending knowledge organization from information retrieval to design judgment. Existing applications of knowledge graphs and RAG in cultural heritage have largely focused on data integration, knowledge retrieval,

and question answering [39, 41]. This study incorporates historical sources, cultural relationships, and user needs into the same evidence selection process. Cultural evidence concerning literati gatherings in the Song dynasty can therefore inform contextual judgments about traditional dress. Each recommendation retains traceable evidence. The five dimensional cultural knowledge graph provides the relational basis for evidence traceability. BERTopic identifies themes in contemporary user perceptions, while AHP converts the three need orientations into interpretable weights. Graph relations, source reliability, and Source-aware MMR regulate how evidence enters the evidence context. Together, these components preserve the links between dress elements and their historical sources, associated people, events, settings, and meanings. User needs affect the priority assigned to evidence. The content and reliability level of the historical sources remain unchanged. This process gives cultural sustainability a specific meaning within the study. The continued circulation of visual elements alone cannot sustain traditional dress as cultural heritage. Its contemporary reuse depends on the preservation of source links, cultural relationships, and usage conditions. Design claims can be traced, their contexts of use can be stated, and their evidential limits can be examined. These features distinguish sustainable reuse from general content generation.

The framework also has several practical applications. Traditional dress designers can use it to examine the sources and cultural contexts associated with specific design elements. In cultural tourism displays and immersive experiences, relational evidence can help coordinate character settings, activity narratives, and spatial design. This may reduce discrepancies between visual representation and cultural content. In art and design education, evidence chains and usage constraints can connect source research, concept development, and proposal justification. Students can then examine how design judgments develop from historical materials.

Several limitations remain. The evidence base focuses on traditional dress associated with literati gatherings in the Song dynasty, and the experimental cases contain relatively clear role and setting constraints. The findings therefore reflect the performance of the framework within this type of design task. Its applicability to other dynasties, dress systems, and cultural heritage objects requires further examination. Future research can expand the evidence base while retaining the current audit criteria and examine the stability of the five dimensional cultural structure across different cultural contexts.

The modeling and evaluation of user needs also have defined boundaries. Online comments were used to identify contemporary perceptions, while AHP weights calibrated three design contexts. These weights do not represent fixed preferences across broader user populations. Expert evaluation compared the relative performance of the methods under predefined criteria, but it did not directly examine prolonged use in professional design practice. Future studies can observe how evidence prompts affect proposal selection, revision, and design judgment in actual design tasks.The experiments used a fixed model configuration and a defined parameter setting. The current results do not establish that this configuration will remain optimal across different models and datasets. Cross model replication and broader parameter sensitivity tests are therefore needed. Further work should also improve the identification of conflicting evidence and the procedures for human review.

6. Conclusions

This study developed and evaluated an NED-GraphRAG framework driven by cultural evidence for the sustainable design transformation of traditional dress associated with literati gatherings in the Song dynasty. Dispersed historical images, texts, artifact reports, and modern studies were organized into a traceable five dimensional cultural

knowledge graph. Contemporary design needs were then incorporated into evidence retrieval and ranking, allowing knowledge of traditional dress to inform specific design judgments.The full framework achieved the highest point estimates for retrieval relevance and stronger results for fit with need orientation and the verifiability of generated outputs. External evidence retrieval strengthened the correspondence between generated content and historical materials, while the graph structure improved the organization of cultural relationships. User perception modeling and AHP need weights adjusted evidence priorities for specific design contexts. The ablation results further showed that AHP need weights, graph relations, source reliability, and Source-aware MMR affected fit with need orientation, ranking quality, evidence quality, and source diversity, respectively. Traditional dress generation therefore requires a balance among relevance, cultural relationships, and source control.The framework extends cultural evidence concerning literati gatherings in the Song dynasty from information organization to design judgment. Dress elements can enter contemporary design transformation while retaining their historical sources and cultural relationships. The resulting proposals remain open to verification. The framework can support traditional dress design and can also inform content organization and proposal development in cultural tourism displays, immersive experiences, and art and design education. Future research can evaluate its applicability across broader dress systems and professional design tasks. Overall, this study offers a practical route for the continued interpretation, transformation, and use of traditional dress culture through traceable cultural evidence.

Author Contributions: Conceptualization, C.L.; methodology, C.L.; software, C.L.; validation, C.L.; formal analysis, C.L.; investigation, C.L.; resources, C.L.; data curation, C.L.; writing—original draft preparation, C.L.; writing—review and editing, C.L.; visualization, C.L.; supervision, P.C.; project administration, P.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical review and approval were waived by the College of Fashion and Design, Donghua University, for this study because it involved non-medical, non-interventional research with minimal risk to participants. No ethics approval number was issued.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank the experts and participants who contributed to the evaluation of this study. During the preparation of this manuscript, the authors used ChatGPT (OpenAI) to refine language and optimize expression. The authors reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ding, Q.-K. and H.-E. Liang, *Digital restoration and reconstruction of heritage clothing: a review.* Heritage Science, 2024. **12**(1): p. 1-21.

2. Chunlan, Y., T.S.M. Tengku Wook, and F. Rosdi, *Advancing cultural heritage: a decadal review of digital transformation in Chinese museums.* NPJ Heritage Science, 2025. **13**(1): p. 189.
3. Xu, J., et al., *A review of the development and application of generative technology in digital museums.* npj Heritage Science, 2025. **13**(1): p. 589.
4. Liu, J., et al., *Preserving China's Intangible Cultural Heritage through AR/VR storytelling, social media, and AI narratives.* npj Heritage Science, 2025. **13**(1): p. 685.
5. Wang, H., et al., *Artificial Intelligence for Sustainable Cultural Heritage: Practical Guidelines and Case-Based Evidence.* Sustainability, 2025. **17**(20): p. 9192.
6. Fu, Y., K. Shi, and L. Xi, *Artificial intelligence and machine learning in the preservation and innovation of intangible cultural heritage: Ethical considerations and design frameworks.* Digital Scholarship in the Humanities, 2025. **40**(2): p. 487-508.
7. Wang, P., et al., *Graph-human: an intelligent question–answering digital human model enhanced by automatically constructed knowledge graph for cultural and tourism applications: P. Wang et al.* Knowledge and Information Systems, 2026. **68**(1): p. 141.
8. Zhu, X., et al. *Knowledge graph-guided retrieval augmented generation*.
9. Zong, Y. and S.-H. Tsaur, *Destination experiencescape for Hanfu tourism: An exploratory study.* Journal of China Tourism Research, 2024. **20**(1): p. 144-166.
10. He, Y., et al., *Hanfu: a systematic literature review integrating cultural tourism, digital technology, and national identity.* Frontiers in Communication, 2026. **11**: p. 1777501.
11. Niu, X., X. Wang, and X. Bi, *Mapping the nexus of sustainability and cultural heritage: A systematic review and empirical analysis.* npj Heritage Science, 2025. **13**(1): p. 471.
12. Paolanti, M., E. Frontoni, and R. Pierdicca, *Towards trustworthy AI in cultural heritage.* npj Heritage Science, 2026. **14**(1): p. 131.
13. Van Vijle, A., P. Hacıgüzeller, and G. Van der Snickt, *Machine learning for painting conservation: a state-of-the-art review.* npj Heritage Science, 2025. **13**(1): p. 437.
14. Zhan, J., et al., *Research on computer vision in intelligent damage monitoring of heritage conservation: the case of Yungang Cave Paintings.* npj Heritage Science, 2025. **13**(1): p. 50.
15. Wu, X., Y. Yu, and Y. Li, *A damage detection network for ancient murals via multi-scale boundary and region feature fusion.* npj Heritage Science, 2025. **13**(1): p. 82.
16. Wu, C., J. Tang, and L. Tong *Generative AI for Cultural Heritage: Shanghai Revolutionary Culture Digital Design Based on the SD–LoRA Model*. Applied Sciences, 2026. **16**, 4427 DOI: 10.3390/app16094427.
17. Lee, D. and E. Palmer, *Prompt engineering in higher education: a systematic review to help inform curricula.* International Journal of Educational Technology in Higher Education, 2025. **22**(1): p. 7.
18. Li, C.J., et al., *Generative AI-supported progressive prompting for professional training: Effects on learning achievement, critical thinking, and cognitive load.* British Journal of Educational Technology, 2025. **56**(6): p. 2550-2572.
19. Fang, C., et al., *Generative AI-enhanced human-AI collaborative conceptual design: A systematic literature review.* Design Studies, 2025. **97**: p. 101300.
20. Wang, S.-F. and C.-C. Chen, *Revitalizing Intangible cultural heritage via derivative design: A focus on chinese woodblock printing.* PloS one, 2025. **20**(4): p. e0318807.
21. Pan, S., et al., *Constructing a sustainable evaluation framework for AIGC technology in Yixing Zisha pottery: balancing heritage preservation and innovation.* Sustainability, 2025. **17**(3): p. 910.
22. Lai, S., Y. Tian, and Q. Zhang, *The impact of AI-generated technologies-driven digital cultural heritage platforms on users' offline cultural participation intentions.* npj Heritage Science, 2025. **13**(1): p. 574.
23. Arzomand, K., T. Kalganova, and M. Rustell, *HARF: A human–AI collaborative framework for cultural heritage reconstruction with expert-guided multi-platform generative AI and systematic prompt engineering.* Digital Applications in Archaeology and Cultural Heritage, 2026. **42**: p. e00554.

24. Dongjiao, Z., Z. Yunlong, and J. Xu, *Research on user satisfaction with AIGC assisted museum scenario design.* Scientific Reports, 2025.
25. Pan, S., et al. *Constructing a Sustainable Evaluation Framework for AIGC Technology in Yixing Zisha Pottery: Balancing Heritage Preservation and Innovation*. Sustainability, 2025. **17**, 910 DOI: 10.3390/su17030910.
26. Li, X., J. Lin, and X. Zhang *Dynamic Transmission and Innovative Transformation of Cultural Heritage: Generative Artificial Intelligence Practices Based on Cultural Cognitive Models*. Applied Sciences, 2025. **15**, 12651 DOI: 10.3390/app152312651.
27. Felicetti, A., A. Himmiche, and M. Somenzi *Knowledge Graphs and Artificial Intelligence for the Implementation of Cognitive Heritage Digital Twins*. Applied Sciences, 2025. **15**, 10061 DOI: 10.3390/app151810061.
28. Huang, Y.Y., et al., *Using knowledge graphs and deep learning algorithms to enhance digital cultural heritage management.* Heritage Science, 2023. **11**(1): p. 1-26.
29. Bai, B. and W. Hou, *The application of knowledge graphs in the Chinese cultural field: the ancient capital culture of Beijing.* Heritage Science, 2023. **11**(1): p. 1-17.
30. Cui, Y., et al., *Linking past insights with contemporary understanding: an ontological and knowledge graph approach to the transmission of ancient Chinese classics.* Heritage Science, 2024. **12**(1): p. 1-18.
31. Liang, W., et al., *A survey of multi-modal knowledge graphs: Technologies and trends.* ACM Computing Surveys, 2024. **56**(11): p. 1-41.
32. Fan, T., H. Wang, and T. Hodel, *Multimodal knowledge graph construction of Chinese traditional operas and sentiment and genre recognition.* Journal of Cultural Heritage, 2023. **62**: p. 32-44.
33. Fan, T., H. Wang, and T. Hodel, *CICHMKG: a large-scale and comprehensive Chinese intangible cultural heritage multimodal knowledge graph.* Heritage Science, 2023. **11**(1): p. 1-18.
34. Wan, J., et al., *WuMKG: a Chinese painting and calligraphy multimodal knowledge graph.* Heritage Science, 2024. **12**(1): p. 1-18.
35. Gao, F., et al. *KA-RAG: Integrating Knowledge Graphs and Agentic Retrieval-Augmented Generation for an Intelligent Educational Question-Answering Model*. Applied Sciences, 2025. **15**, 12547 DOI: 10.3390/app152312547.
36. Huang, Y. and J.X. Huang, *A survey on retrieval-augmented text generation for large language models.* ACM Computing Surveys, 2026. **58**(12): p. 1-38.
37. Peng, B., et al., *Graph retrieval-augmented generation: A survey.* ACM Transactions on Information Systems, 2025. **44**(2): p. 1-52.
38. Knollmeyer, S., O. Caymazer, and D. Grossmann, *Document graphrag: knowledge graph enhanced retrieval augmented generation for document question answering within the manufacturing domain.* Electronics, 2025. **14**(11): p. 2102.
39. Xu, L., et al., *Nanjing Yunjin intelligent question-answering system based on knowledge graphs and retrieval augmented generation technology.* Heritage Science, 2024. **12**(1): p. 1-23.
40. Xu, J., et al., *ChatTf: a knowledge graph-enhanced intelligent Q&A system for mitigating factuality hallucinations in traditional folklore.* IEEE Access, 2024. **12**: p. 162638-162650.
41. Yuan, H., et al., *Knowledge graph-based intelligent question answering system for ancient Chinese costume heritage.* npj Heritage Science, 2025. **13**(1): p. 198.
42. Bourdieu, P., *Distinction: A social critique of the judgement of taste*, in *Social Stratification, Class, Race, and Gender in Sociological Perspective, Second Edition*. 2019, Routledge. p. 499-525.
43. Bourdieu, P., *The field of cultural production, or: The economic world reversed.* Poetics, 1983. **12**(4-5): p. 311-356.
44. Husu, H.-M., *Rethinking incumbency: Utilising Bourdieu's field, capital, and habitus to explain energy transitions.* Energy Research & Social Science, 2022. **93**: p. 102825.
45. Dörry, S. and C. Schulz, *Creating Low-Carbon Economies: Probing Transition Dynamics through the Lens of Field Theory.* Tijdschrift voor economische en sociale geografie, 2024. **115**(5): p. 660-673.
46. Geerts, R., F. Vandermoere, and S. Oosterlynck, *Culture and green tastes. A sociological analysis of the relationship between cultural engagement and environmental practices.* Poetics, 2023. **97**: p. 101768.

47. Dicks, B., *The habitus of heritage: A discussion of Bourdieu's ideas for visitor studies in heritage and museums.* Museum & Society, 2016. **14**(1): p. 52-64.
48. Evans, S. and R. Whiting, *Everyday constructions of class-based inequality: field effects in UK museum work.* Sociology, 2025. **59**(2): p. 363-379.
49. Cun, S., et al., *Beyond Information: A Dual-Path Strategy for Sustainable Digital-Cultural-Heritage Management Driven by Affective Experience.* Sustainability, 2026. **18**(2): p. 699.
50. Meng, L. and Z.-Y. Zhu, *How Does Digital Experience of Cultural Heritage Transform into Sustained Behavioral Intention? Assessing Perceived Value and Place Attachment Mechanisms Based on Value Adoption Model.* Sustainability, 2026. **18**(3): p. 1470.
51. Zhong, L., et al., *Who wears Hanfu clothing when they travel? Experience-based cultural tourism market segmentation–the case of Xi'an, China.* Current Issues in Tourism, 2023.
52. Zong, Y., J.S. Chen, and S.-H. Tsaur, *Destination experiencescape for Hanfu tourism: Scale development and validation.* Journal of Vacation Marketing, 2025. **31**(2): p. 314-334.
53. Hou, L., et al., *Sustainable design of museum bronze cultural-creative products through an integrated Kano-AHP-TOPSIS approach.* Discover Sustainability, 2025. **6**(1): p. 1212.
54. Wang, Z., et al., *An Integrated KANO–AHP–DEMATEL–VIKOR Framework for Sustainable Design Decision Evaluation of Museum Cultural and Creative Products.* Sustainability, 2025. **17**(22): p. 10328.
55. Hong, C., et al., *A Kansei-inspired product design approach for the Liao Dynasty Pagoda: bridging cultural heritage and physiological signals.* Journal of Engineering Design, 2026. **37**(1): p. 94-123.
56. Hu, P., C. Chen, and Z. Wang, *Innovative Design of Contemporary Children's Wear Based on the Cultural Gene Map of Children's Dress in the Song Dynasty.* Journal of Clothing Research, 2025. **10**(4): p. 347-357.
57. Liu, C. and Q. Lei, *Gene Map Construction and Design Translation of Yurt Culture from the Perspective of Cultural Field.* Packaging Engineering, 2023. **44**(6): p. 286-301+310.
58. McInnes, L., J. Healy, and N. Saul, *Großberger, LUMAP Uniform manifold approximation and projection.* J. Open. Source. Softw, 2018. **3**: p. 861.
59. McInnes, L., J. Healy, and S. Astels, *hdbscan: Hierarchical density based clustering.* J. Open Source Softw., 2017. **2**(11): p. 205.
60. Saaty, T.L., *How to make a decision: the analytic hierarchy process.* European journal of operational research, 1990. **48**(1): p. 9-26.