CultiVerse: Towards Cross-Cultural Understanding for Paintings with Large Language Model

Wei Zhang 💿, Wong Kam-Kwai 💿, Biying Xu, Yiwen Ren, Yuhuai Li, Minfeng Zhu, Yingchaojie Feng, and Wei Chen



Fig. 1: The interface of CultiVerse offers a streamlined approach to exploring and appreciating paints for enhanced cross-cultural understanding. It starts with the source culture extraction (A), which provides accurate information about the TCP for identifying and associating visual elements. Next, the culture exchange (B) shows elements and symbols across cultures, aiming to align the disparities in diverse cultural norms and provide LLM-generated content to support multimodal understanding. Lastly, the target culture extrapolation (C) encourages curiosity-led exploration while ensuring adherence to cultural norms and appropriateness.

Abstract— The integration of new technology with cultural studies enhances our understanding of cultural heritage but often struggles to connect with diverse audiences. It is challenging to align personal interpretations with the intended meanings across different cultures. Our study investigates the important factors in appreciating art from a cross-cultural perspective. We explore the application of Large Language Models (LLMs) to bridge the cultural and language barriers in understanding Traditional Chinese Paintings (TCPs). We present CultiVerse, a visual analytics system that utilizes LLMs within a mixed-initiative framework, enhancing interpretative appreciation of TCP in a cross-cultural dialogue. CultiVerse addresses the challenge of translating the nuanced symbolism in art, which involves interpreting complex cultural contexts, aligning cross-cultural symbols, and validating cultural acceptance. CultiVerse integrates an interactive interface with the analytical capability of LLMs to explore a curated TCP dataset, facilitating the analysis of multifaceted symbolic meanings and the exploration of cross-cultural serendipitous discoveries. Empirical evaluations affirm that CultiVerse significantly improves cross-cultural understanding, offering deeper insights and engaging art appreciation.

Index Terms—Cross-cultural understanding, Traditional Chinese painting, Large language model, Mixed-initiative workflow

1 INTRODUCTION

- W. Zhang, B. Yu, Y. Ren, Y. Li, M. Zhu, Y. Feng, and W. Chen are with the State Key Lab of CAD&CG, Zhejiang University. Wei Chen (corresponding author) is also with the Laboratory of Art and Archaeology Image (Zhejiang University), Ministry of Education, China. E-mail: {zw_vixian, biyingx, renyiwen, yuhuaili, minfeng_zhu, fycj, chenvis}@zju.edu.cn.
- W. Kam-Kwai is with Hong Kong University of Science and Technology and Georgia Institute of Technology. Email: kkwongar@connect.ust.hk.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx The intersection of technology and cultural studies has significantly reshaped the analysis of historical narratives through modern perspectives [63, 64, 66]. These advanced studies, rich in historical insight, facilitate in-depth exploration of cultural heritages [16, 53, 65]. Yet, the intricate design of such tools is often tailored for experts, limiting their accessibility to reach a broader audience who may lack essential prior knowledge. The issue is exacerbated for individuals from different cultural backgrounds. They may not be aware of the necessary historical and cultural artifacts. Without adequate guidance, there is a risk that the intended implications could be misinterpreted or distorted with biased perceptions. This scenario motivates the pressing need to create systems that make insightful cultural analyses accessible to a broader audience regardless of their prior knowledge or cultural backgrounds.

In this context, cross-cultural understanding is crucial in fostering the ability to recognize, interpret, and appropriately respond to cultural phenomena (e.g., language, customs, social norms, and art) that seem unfamiliar or exotic due to cultural differences. This understanding is essential for nurturing effective intercultural communication and enhancing multicultural awareness and sensitivity. Our study leverages Traditional Chinese Paintings (TCP) as a case in point to deepen cross-cultural understanding. The visual aspect of these paintings transcends spoken and written language barriers, offering a universally accessible medium to explore and understand the underlying values of another culture. By engaging with and interpreting artworks from varied cultural backgrounds, individuals can appreciate different cultures' distinctive perspectives and ways of life, laying a foundation for mutual understanding. Fostering cross-cultural understanding not only enriches individual perspectives but also contributes to the promotion of broader cultural exchange and harmony.

We explore using Large Language Models (LLMs) to bridge cultural and language barriers for cross-cultural understanding. Pretrained on a diverse corpus of high-quality texts from different cultures, these models are adept at analyzing and contextualizing complex information. They also synergize with generative image models, enabling image creation in diverse artistic styles upon request, which aids in visualizing complex concepts that are difficult to articulate with words alone. Utilizing LLMs offers unprecedented opportunities to enhance the cross-cultural understanding of paintings, making painting appreciation more inclusive to people from different cultural backgrounds.

However, in developing system prototypes, we have identified three key challenges in harnessing LLMs for such purposes. We illustrate them using the lotus flower as an example. In TCP, it symbolizes purity and spiritual enlightenment when portrayed alone [42] (Fig. 2A), while its meaning shifts to represent a harmonious love relationship when associated with mandarin ducks [47] (Fig. 2B). This shift exemplifies the challenge of contextual interpretation within the multifaceted symbolic meanings. In contrast, the lotus suggests a different connotation of divinity and creation in Hindu art [22] (Fig. 2C), and signifies rebirth and restoration in Ancient Egyptian culture [50]. This highlights the challenge of cross-cultural symbol alignment. Furthermore, the dragon provides another illustrative example that is respected in Chinese culture as a symbol of strength and prosperity, but is often associated with threat and malevolence in Christian culture [61]. This divergence highlights the difficulty in validating variations in different cultural customs. Despite their proficiency in processing textual information, LLMs exhibit limitations in accurately translating the rich visual narratives of cultural arts. Human intervention is required to guide the sense-making process and prevent cultural misinterpretations or oversimplifications.

We introduce CultiVerse [4], a visual analytics system that harnesses LLMs in a mixed-initiative framework, designed for novices to enhance the interpretive engagement with cultural artifacts in a cross-cultural dialogue. We first characterize the domain concept of "cultural norm" and curate a corresponding TCP dataset. This semi-automatically annotated dataset captures the broad spectrum of symbolic meanings associated with various elements depicted in TCP. Then, we design an interactive interface that visualizes the multifaceted symbolic meanings and intercultural connections, enabling users to filter and analyze the cultural norms effectively. The interface seamlessly integrates LLMs to augment the cross-cultural dialogue with enriched contextual understanding, align cross-cultural interpretations, and propose alternative associative ideas to inspire serendipitous discoveries. Furthermore, we have incorporated pre-configured structured prompts to assist users in efficiently validating the cultural acceptability of their exploratory findings. CultiVerse's mixed-initiative workflow progressively guides both the user and the LLM towards generating more refined prompts and results, thus optimizing the interpretive accuracy. Through comparative studies and user interviews, we have demonstrated that CultiVerse significantly improves the efficiency of cross-cultural understanding and yields more varied insights than conventional methods. Our system not only streamlines the analytical process but also deepens the level of interpretative exploration, providing transformative perspectives into the interplay of culture and art. This study has the following contributions:



Fig. 2: The lotus flower carries diverse meanings in different cultural contexts. (A) The *Lotus Painting*" by Xu Wei symbolizes nobility [42]. (B) The *Lotus Pond and Mandarin Ducks*" by Shen Quan expresses a harmonious love relationship [47]. (C) Painting from the Bhagavata Purana, a Hindu manuscript, describes creation [22].

- We characterize the problem of cross-cultural understanding for TCP and curate a TCP Cultural Norm Dataset that captures the diverse symbolic meanings of elements.
- We develop CultiVerse based on the proposed mixed-initiative workflow, leveraging visual analytic approaches and LLMs.
- We conduct comparative user studies and interviews to demonstrate the effectiveness and usefulness of CultiVerse.

2 RELATED WORKS

2.1 Visual analytics for cultural understanding

Cultural understanding is a complex process that involves recognizing, appreciating, and interpreting specific cultural phenomena [40]. Cultural artifacts such as ancient books [64], ancient pottery [60], and paintings [14] play a crucial role in cultural understanding. Researchers use digital tools to uncover the cultural phenomena hidden behind these artifacts. The rapid development of digital humanities in recent years demonstrates that visual analytics is an effective method to enhance cultural understanding and improve its efficiency [11]. For example, Zhang *et al.* [65] used visual analytics to delve deeper into the socialcultural context of the Song Dynasty through poetry. LiberRoad [21] used ancient books as a medium to visualize the international dissemination of Chinese classical literature, promoting cultural exchange and integration between China and Japan. ScrollTimes [63] facilitated art historians in tracing the provenance of paintings by extracting elements from these paintings and conducting visual associative analyses.

However, these existing works often target a niche audience of experts for in-depth historical analysis, requiring extensive domain knowledge and often overlooking the broader goal of making cultural art and phenomena more accessible to the general public. To bridge this gap, we have designed and developed a visual analytic system, CultiVerse, to bring insights related to cultural symbols to a general audience, particularly those from diverse cultural backgrounds and those interested in learning about other cultures.

2.2 Facilitating cross-cultural understanding

Cross-cultural understanding extends the definition of cultural understanding to include language, customs, social norms, and artistic expressions that might be unfamiliar due to cultural discrepancies [29, 40, 41]. This understanding is crucial for effective communication and interaction in our increasingly globalized world [44].

Amanda [23] pointed out that in the absence of a common language, painting can serve as vital modes of communication for understanding other cultures. Liu [31] discussed the methodology used in the author's paintings to visualize cross-culturalism. Emilie [33] investigated a cross-cultural experiment that combined classical music and soundscapes with painting to enhance intercultural competence. Zhao *et al.* [69] designed a tablet-based authoring tool to support cross-cultural appreciation of TCP through interactive experiences. Similarly, Pan *et al.* [37] examined the aesthetic principles in TCP and Western cultural active factors, aiming to deconstruct and distill cultural elements for modern product design processes. However, these works mainly focus on appreciating the artistic style of paintings, lacking in-depth translation and interpretation of the cultural meanings conveyed by the paintings. Our work, through meticulous requirement research, outlines a detailed analytical process for cross-cultural understanding based on paintings for audiences from diverse cultural backgrounds.

Reducing misunderstandings is also a crucial issue in cross-cultural understanding [10, 25]. Taracenko et al. [49] suggest that fundamental educational training is necessary, such as enhancing basic knowledge of different countries, and learning about relevant languages and cultural characteristics. However, this process might be challenging and require significant effort. The development of LLMs provides a technical foundation for minimizing misunderstandings in cross-cultural understanding. Regarding data sources, Liu et al. [32] developed a Cross-Cultural Understanding Benchmark (CCUB) dataset aimed at achieving multicultural inclusivity and representation in generated images. In terms of exploration processes, NormSage [20] offers a framework for discovering dialogue-driven, multilingual, and multicultural norms, based on language model prompting and self-validation. Inspired by these prompt engineering approaches, we explored the incorporation of visual elements in TCPs with LLMs to facilitate cross-cultural understanding. We curate an expert-annotated TCP dataset based on the concept of cultural norm to provide accurate information for LLM generation.

2.3 Incorporating with Large Language Models

After scaling the parameter capacity of pre-trained language models [28], researchers found those LLMs showed surprising performance (called emergent abilities [54]) in solving diverse, sophisticated tasks [12, 13, 51, 70]. With their outstanding instruction-following [62], in-context learning [12, 52], and reasoning [55] abilities, LLMs have been leveraged to empower various domain applications. These include a few attempts [26, 58] to apply LLMs on cultural and social sciences, which mainly focus on studying LLMs' inherent social bias [7, 35] and modeling the sociocultural norms [38, 39, 45].

Unlike previous works, this paper is the first to adopt the powerful GPT-4 [6] model as the inference backend of our visual analytic system CultiVerse. We provide empirical evidence to show that LLMs' profound world knowledge can benefit cross-cultural understanding. Our carefully designed pipeline enhances the usability and authenticity of culture translation for the public interest.

3 DESIGN REQUIREMENT ANALYSIS

We have collaborated with domain experts to inform our system design, which aims to engage a broader audience with TCP through a cross-cultural dialogue. We interviewed four professors, two in TCP and two in cultural translation. E1 and E2 have 15 and 30 years of TCP research experience. E3 has dedicated 15 years to studying Chinese-Japanese literature and has a deep understanding of TCP. E4 focuses on comparative educational research between China and the UK. None of them is a co-author. Through semi-structured, one-on-one interviews, we explored a hypothetical scenario where experts assist non-natives in understanding a TCP. Then, we developed a prototype system incorporating ChatGPT-4 Turbo [2] to re-examine the scenario. All experts have recognized LLM's ability to bridge cultural and language barriers for cross-cultural understanding. However, the inherent complexity of TCP's cultural connotations presents challenges that primitive LLM applications cannot adequately address. Through this iterative process, we collected expert feedback on strategies for promoting art appreciation and the challenges of using LLMs for cross-cultural understanding.

3.1 Painting analysis process

E1 and E2 suggest three levels of comprehension (L1-L3) that progressively deepen an audience's appreciation of TCP, requiring increasing



Fig. 3: The design requirements are inspired by TCP comprehension levels (A-C) and the literature translation process (D).

cultural understanding and artistic engagement.

L1. Visual level focuses on the immediate observation and the surface appreciation of art's visual elements. For instance, Shen Zhenlin's *"Elderly in Spring Painting"* (Fig. 3A) vividly portrays a cat, butterflies, and chrysanthemums with delicate brushwork. This initial stage emphasizes aesthetics appreciation, including recognizing elements within the TCP and appreciating the drawing techniques used [19].

L2. Symbol level involves identifying relationships between elements and understanding why these elements are presented in specific ways. It explores the symbolic meanings behind the co-occurrence of certain elements [46, 57]. For example, butterflies generally symbolize love in Chinese culture, but when depicted with cats, as in Shen Zhenlin's painting (Fig. 3B), they suggest longevity. This interpretation originates from the homophonic play between "猫蝶 māodié [cats and butterfly]" and "耄耋 màodié [being from 80 to 89 years old]," illustrating symbolic associations embedded within the artwork.

L3. Custom level seeks to understand the painter's intent and the cultural phenomena represented by the symbols [8, 56]. This deepest level of appreciation requires a comprehensive understanding of the cultural context to connect the artwork to broader cultural practices and beliefs. For example, presenting the painting named "*Elderly in Spring Painting*" (Fig. 3C) as a gift reflects societal values placed on a joyful, fulfilling, and vibrant old age due to respect and care for elders.

3.2 Design requirement for cross-cultural understanding

We gain inspiration from E3 and E4's experience in the literature translation process, which includes source culture analysis, culture adaptation, and target culture assimilation [9], for the design goals of cross-cultural understanding. We further contextualize them with TCP comprehension levels and summarize the design requirements.

G1. Source culture analysis. The foundation of accurate cultural translation lies in a deep understanding of the source culture. E4 warns against the risks of biased translations stemming from insufficient understanding of the source material, akin to the notion "garbage in, garbage out," which can significantly increase the chance of invalid LLMs' responses. To mitigate this, LLMs should be provided with accurate information about the three levels of TCP comprehension, ensuring a more accurate and culturally sensitive translation process.

R1. Identifying visual elements. From natural to mythological elements, TCPs are rich in artistic expressions and drawing techniques that differ from other artistic styles (*e.g.*, oblique projection [67]). Developing models trained with diverse representative samples is crucial for accurate element identification at the visual level (L1).

R2. Visualizing possible associations between elements. The tradition of replicating classics in TCPs (called "to transmit models by drawing" [67]) suggests that widely accepted cultural symbols and elements are repeatedly drawn. Therefore, presenting recurring elements among different paintings can shed light on the potential connections within their original context [27]. Moreover, transparently embedding authoritative historical and cultural resources in the system offers users and LLMs a trusted foundation for generating content with cultural insights [18]. These help understand the symbol level (L2).

G2. Culture adaptation. Adapting elements and symbols from the

source culture to fit the target culture's norms requires a delicate balancing effort, often challenged by disparities in symbolic and customary meanings. Gathering relevant information from both cultures is important to refine and adjust the adaptations. This iterative process enhances user engagement by offering familiarity and new insights.

R3. Specifying the target culture. LLM responses can accommodate diverse user backgrounds. To bridge language and cultural barriers, specifying users' cultural contexts and offering feedback on cultural relevance facilitates more meaningful and personalized translations.

R4. Aligning cross-cultural symbols. The challenge of aligning symbolic meanings across cultures stems from L1-mismatches (where different elements represent the same symbol) and L2-mismatches (where the same element represents different symbols). The system should enable users to visually compare elements and symbols from both cultures to identify parallels or divergences in symbolic meaning.

R5. Supporting multimodal understanding. E4 notes that illustrations and other visual aids are common strategies for literature translation. They depict relatable scenes or realistic source culture objects, which are particularly useful for L3-mismatches where customs diverge significantly between cultures. Multimodal translations provide an enriched and engaging context for cross-cultural understanding.

G3. Target culture assimilation. Integrating aligned symbols and customs into the target culture requires a process of creative exploration that honors cultural nuances. Employing LLMs to generate content rich in cultural relevance, such as narratives that find echo in both the source and target cultures, showcases a profound cross-cultural understanding. Nevertheless, such content should maintain cultural appropriateness.

R6. Supporting creative exploration. By offering structured guidance around the users' natural curiosity, the system could spark a desire to explore TCPs across different cultural contexts. This curiosity-led approach facilitates users to form a personal connection with the materials. Encouraging questions and self-directed discovery allows for a richer, more immersive cross-cultural experience.

R7. Validating norm acceptability. E3 and E4 found that some LLM-generated exploration results may deviate from the source culture's values. Since a general audience often lacks awareness of technical limitations such as content hallucination [68], taking the result without validation can risk cultural biases. The system should have a safeguarding mechanism for validating the cultural appropriateness of content and preserving the integrity of cross-cultural understanding.

4 Source cultural NORM DATASET CONSTRUCTION

This section describes the *TCP Cultural Norm Dataset (TCP-CND)* for source culture understanding (G1). We adopt a semi-automated approach for efficient, accurate dataset construction. Four TCP domain experts are invited to annotate TCP data, aiding in selection, optimization, and annotation. We select English as the primary language for the dataset to accommodate its wide adoption in the global population and LLMs' language ability, reducing the likelihood of misunderstandings in further cross-cultural translation processes.

4.1 Cultural norm definition

Based on our requirement analysis and cultural studies [43], we define the data structure of *Cultural Norm* (*CN*) as follows:

$$CN = (Element - Rhetoric - Symbol - Custom - Emotion)$$
(1)

where *Element* refers to the objects that appear in the painting. *Rhetoric* denotes the technique used to create a symbolic association between the "element" and the "symbol." *Symbol* refers to the symbolic meaning that the "element" is associated with, using a specific "rhetoric" technique. *Custom* corresponds to the cultural, societal, or traditional practice or belief associated with the element and its symbol. *Emotion* describes the feeling or sentiment associated with or evoked by using the symbol within the specific cultural context.

4.2 Data collection and annotation

We collect TCP information from two open datasets, the National Palace Museum (NPM) [34] and the Chinese Iconography Thesaurus (CIT) [3]. Each painting contains the following data information:

Painting image and background information. Under the guidance of experts, we carefully selected a collection of paintings, such as flower and bird paintings, imbued with diverse implications that are more conducive to public users. We excluded landscape paintings and portraits with complicated historical backgrounds. Ultimately, we obtained 534 pieces of paintings. Each painting is characterized by the artist, dynasty, medium, dimensions, and location.

Painting element. The CIT dataset establishes a thesaurus for TCP, including general elements (e.g., plants, animals, characters, and events) and proper nouns (e.g., locations, myths, and literature). However, some elements, such as "looking upward (bird)" representing animal forms, offer limited assistance in cross-cultural understanding. We collaborated with experts to filter these elements. Finally, we selected a total of 226 elements (plant 94, animal 86, fruit 16, other 13, composite elements 17). However, the lack of instance-level annotations (i.e., bounding boxes and categories) hinder the in-depth exploration of the painting. We adopted a human-in-the-loop labeling pipeline to guarantee high data quality at reasonable costs. We utilized GroundingDINO [30], a prompt-based object detection model, to obtain the bounding boxes of visual elements. Due to natural degradation, some TCPs exhibit color fading [48] and contain elements in unique art forms, undermining automatic recognition's reliability and fall short of our annotation standards. We invite four experts to refine and augment the annotations. Each annotation is thoroughly verified by two experts for accuracy and consistency in (Image - Element) alignment (R1).

4.3 Norm extraction

We cooperate with experts to extract, organize, optimize, and enrich norms from the "Chinese Cultural Symbol Dictionary" [15]. Ultimately, we obtained 505 cultural norms.

Element, symbol, and custom extraction. The dictionary contains various symbolic meanings of the elements and their origins in Chinese culture. We classify elements into *atomic* and *composite elements*. The composite elements are generated from multiple atomic elements via the *and* logical operator to represent more complicated contexts. For instance, atomic (蜂 fēng [Bee]) & atomic (猴 hóu [Monkey]) indicates the composite (蜂猴 fēnghóu = 封侯 fēnghóu [being ennobled as a marquess]). However, not all of these apply to our study. For instance, elements such as "cold," "constellations," "dust," and "earth" pose challenges in mapping their visual representation through painting. As such, we refine these data and ultimately extract 1453 elements, along with their various symbols and customs (R2).

Rhetoric and emotion identification. To foster a more profound comprehension of the emotions associated with customs, experts annotate each "element - symbol - custom" pair. The emotions can be categorized into three types: "positive," "negative," or "neutral." Experts also annotate the rhetorical techniques employed in the "element - symbol" pairs, elucidating the method of transforming an "element" into a "symbol," such as (蜂猴 fēnghóu - Homophony - 封侯 fēnghóu). We summarize six rhetorical techniques from the paintings and iterate their definitions with E1 and E2. The definitions for each rhetorical technique are as follows:

Iconic: the form of the sign is connected to its meaning.

• Homophony: two words have the same pronunciation but different meanings, origins, or spelling.

• Homophonic pun: a type of pun that exploits the fact that words sound similar but have different meanings.

• Synonym: a word that means exactly or nearly the same as another word.

• Homograph: a word that shares the same form as another word but has a different meaning.

• Satire: a word with a satirical meaning often reflects negative aspects of society.

5 CULTIVERSE

This section describes the workflow, engineered prompts for the components, and interactions between views in CultiVerse (Fig. 4).



Fig. 4: The CultiVerse workflow comprises three steps: source culture extraction (A), culture exchange (B), and target culture extrapolation (C). Initially, user observe element distribution (A1), examine element correlations (A2), and select painting based on exploratory interests (A3), followed by choosing element for analysis (A4). Subsequently, user analyze (B1), generate (B2), and translate (B3) cultural norms in conjunction with LLMs. Finally, users validate (C1) and expand (C2) the translation results.

(A)Image frequency

CRhetoric frequency & type Symbol (ZH / EN) Emotion type



Emotion Type: ©Positive ⊗Negative ©Neutral

Rhetoric Type: Iconic Homophony Homophonic pun Synonym Homograph Satire

Fig. 5: The element selection view highlights the frequency of elements in the dataset (A) and counts of other factors in the cultural norm (B-F).

5.1 User background setting

Initially, users are required to configure background information (G2, R3). This role information will be utilized as a prompt setting, transmitted to the LLM, and integrated throughout the entire analysis workflow. The data structure of *User Background (UB)* is as follows:

$$UB = (Country, Age, Education, FWC, FWTCP, Note)$$
 (2)

where *FWC* and *FWTCP* refer to the user's familiarity with Chinese culture and TCP on a scale of 1 to 5, respectively. *Note* allows users to specify contextual information, such as cultural preferences for anime.

5.2 Source culture extraction component

The source culture extraction component (Fig. 4A) offers an element feature view that illustrates data features based on the TCP-CND (G1, R1, R2). Users can then explore these data features and select a painting that interests them. The chosen element from the painting is subsequently forwarded to the following analysis stages.

5.2.1 Element feature view

The element feature view visualizes all identified elements derived from the dataset (R1). To enhance the understanding of data features, this view illustrates each element's frequency, relevance, and overall distribution across various categories (R2). We encode the elements through distinct channels to differentiate them.

Visualizing elements in colors, sizes, and lines. The extracted features could be either atomic or composite elements (see Sec. 4.3). We use different colors to distinguish the four categories of atomic elements (*i.e.*, \Diamond *Animal*, \Diamond *Plant*, \bigcirc *Fruit*, and \bigotimes *Others*). We encode the number of elements by size to emphasize the data distinction (Fig. 4A1). Zooming in an atomic element reveals its composite elements. The composite elements combine the lines of their corresponding elements. The lines' thickness encodes the number of element's relevance (Fig. 4A2).

5.2.2 Painting selection view

When a user selects an element in the element feature view, a related painting is displayed in the painting selection view (Fig. 4A3). The user can then choose one for further analysis. Element annotations from the TCP-CND, which includes the element bounding boxes and their corresponding labels, are shown in the painting (R1). If there are elements of interest to the user that have not been annotated, we provide a manual annotation tool to assist the user in this task. The newly added annotations are also recorded in the element selection view (Fig. 4A4). The painting's background information is presented on demand for more detailed information. Image interaction tools (*e.g.*, fullscreen, zoom in, zoom out) are provided to facilitate close investigation of visual elements in TCPs.

5.2.3 Element selection view

The element selection view (Fig. 1A4) provides detailed information about the elements of the selected painting (R1). Numbers in Fig. 5A represent the frequency of each image. The related elements of the image are displayed below (Fig. 5B) in Chinese and English. The colors in Fig. 5C indicate the categories of rhetorical types involved with each element, and the numbers denote the frequency of usage of these types. When composite elements are present, they are linked to related atomic elements and enhanced with icons for clearer representation (R2). The elements are responsive to user-annotated elements if they belong to the dataset, facilitating ad-hoc and comprehensive exploration.

5.3 Culture exchange component

Upon selecting an element, users can explore its diverse symbolic meanings in this component by harnessing the capabilities of the LLM (Fig. 4B). The selected cultural norm can be adapted to the target culture for cross-cultural understanding (Fig. 4B1-B3) (R4-R6).

5.3.1 Source culture exploration view

The source culture exploration view (Fig. 1B1) visualizes the source cultural norm and provides multiple ways to explore it (G1). When an element is selected, all related norms are displayed. Fig. 5D1 and D2 present the different symbols in both Chinese and English. The left side illustrates the rhetoric employed in the 'element-symbol' pairs, differentiated by various colors. Beneath each symbol is a custom explanation, which elucidates why the element has been assigned this symbol, presented in both Chinese and English (Fig. 5E). The right side of the symbol indicates the emotion associated with its use within a specific culture, categorized into three types and distinguished by different icons (Fig. 5F).

Explore with question-answering. We utilize the powerful question-answering capabilities of the LLM to allow users to pose queries about any unclear aspects of the source cultural norm. To provide guidance and prompts, we have pre-set three recommended questions. Additionally, users are free to ask their questions. The structure of the question-answering prompt is as follows:

$$QA = LLM(UB, CN, PR \mid FQ)$$
(3)

Where UB denotes the user background in Eq. (2), CN represents the source cultural norm in Eq. (1), PR denotes the prompt recommendation, and FQ signifies the free question. The question-answering process incorporates a memory function, enabling users to continue asking questions based on previous responses. A deletion function is also provided on the right side of the dialogue box, allowing users to remove unnecessary dialogue boxes to conserve viewing space.

Explore with image generation. To enhance users' understanding of the source cultural norms, we incorporate LLM into image generation (IG). It aims to augment the comprehension of source cultural norms visually (R5). To enable the LLM to generate images that align with users' cognition, we designed the following prompt:

$$IG = LLM(UB, CN, T) \tag{4}$$

where T corresponds to the image generation task. Once the image is generated, options for deletion and regeneration are provided to cater to users' requirements for multiple generations.

5.3.2 Culture transfer view

The culture transfer view (Fig. 1B4) integrates the LLM to offer culture translation services to users (R6). The output of LLM is visualized with the same design scheme as the source cultural norm (Fig. 5) to maintain visual coherence. The view also supports Image Generation and Question-Answering to assist users in further exploration.

Multiple conditions. To cater to the diverse culture transfer needs of users, we provide an interactive function, enabling users to freely choose a part or all of the source cultural norm as the conditions (C):

$$C \in source \ culture \{ Element, Rhetoric, Symbol, Custom, Emotion \}$$
 (5)

For question (Q) formulation, users can interactively choose whether on a single part or multiple parts of the target cultural norm:

$$Q \in target \ culture \{ Element, Rhetoric, Symbol, Custom, Emotion \}$$
 (6)

The selected conditions and questions become a part of the prompt:

$$CT = LLM(UB, CN, E, C, Q, OF)$$
(7)

where E refers to the explanation of the definition, such as rhetoric types. *OF* refers to the pre-defined output format for the view.

5.4 Target culture extrapolation view

In this stage, users can leverage LLM to verify the appropriateness of the explored norms and expand on similar norms (Fig. 1D) (G3, R7).

5.4.1 Cultural norm verification

We put an "Appropriate" button, enabling users to obtain a question template to validate cultural appropriateness (R7). The content contained in the template is as described in the previous paragraph. After sending the question to LLM, it can give a judgment based on the target culture setting and current translation path, and analysis about it.

5.4.2 Cultural norm expansion

Also, we provide an "Inference" button for extending similar norms in a way beyond the workflow (R6), drawing inferences from one instance as a looser and freer supplement to the previous translation method. Al Autumn egret in hibiscus flowe



Fig. 6: Case 2 explores (A1) "Autumn Egret in Hibiscus Flower Landscape." (A2) Analyzing compositional elements to (A3) understand their significance in Chinese culture; and (B1) using interactive annotations to (B2) explore new elements (B3) in Indonesian culture.

6 EVALUATION

We conducted a user study to evaluate how CultiVerse assists a broader audience in achieving cross-cultural understanding through TCP. We invited ten participants (P1-P10) from diverse cultural backgrounds to assess the effectiveness and usability of CultiVerse through comparative studies and free exploration. The explorations by P3 and P10 are presented as the case studies, showing how CultiVerse facilitates crosscultural understanding when users have knowledge in source culture (Sec. 6.1) and target culture (Sec. 6.2). We use labels (1-10) to denote the new insights gained during their exploration. We summarize user feedback on the system design, workflow, and exploration patterns.

6.1 Exploring diverse cultures

P3 (nicknamed Kay) is from China and, working as a consultant, has a deep appreciation for Japanese culture. She holds a Japanese-Language Proficiency Test N2 certificate and has traveled to Japan numerous times. In the target culture panel, she sets her sights on Japan, embarking on an in-depth exploration and study.

Source culture extraction. Initially, Kay closely looked at the distribution of the *element feature view* (Fig. 1A1). ① Among the word cloud, "monkey" popped out and interested her, reminding her of her experiences observing monkeys up close at the Jigokudani Monkey Park in Japan. This triggered her curiosity about monkeys' representation and potential symbolism within TCPs.

The *painting selection view* (Fig. 1A2) displayed all eight paintings from the dataset that included the monkey element. ⁽²⁾ Kay selected and obtained background information about the "Monkey Painting" by Yan Hui in the Mongol Empire (1271-1368) (G1). The two monkeys and a swarm of bees were automatically identified in the painting (L1, R1). ⁽³⁾ She marveled at the intricate depiction of bees, stating that she would not have noticed them in the painting without the annotation. She enlarged the painting for a close examination. She thought the "leaf" might be a missed element, and annotated it herself (Fig. 1A3).

Kay navigated to the *element selection view* (Fig. 1A4) for the elements' features. Through the statistics in the top-right corner, () she noticed that the monkey is associated with three symbols and rhetorical techniques (L2), and forms a composite element (R2) with the bees, termed "bee & monkey." () The newly annotated element "leaf" did not carry symbolic meanings. She clicked on the "bee & monkey." for further analysis in the *culture exchange component* (Fig. 1B).

Culture exchange. In the *source culture exploration view* (Fig. 1B1), "bee & monkey" was explained as symbolizing "being ennobled as a marquess" because the two phrases are associated with homophony. They frequently appear together in paintings, embodying aspirations for upward mobility [53] (Fig. 1B2). (6) Kay was surprised about the implications of "bee & monkey." She wondered if other cultural contexts exist for the connection between the monkey and the marquess



Fig. 7: In the second exploration phase, an unfamiliar element (A) was analyzed to understand its symbols in Chinese culture (B, C), with image generation enhancing visual understanding (D). The second symbol was then translated into the Indonesian cultural context (E), further improving comprehension through image generation (E1,E2).

(R6). While the typical cultural research could take pages of reading, she simply conversed with the LLM and got a satisfactory result. ⑦ Monkeys are symbols of intelligence and wit; Therefore, drawing one may imply a tribute to the wisdom and mastermind of nobles (L3). Also, she prompted the LLM to generate an image related to an ancient Chinese marquess (Fig. 1B3). This visual aid spelled out the hidden meaning in the painting for her (R5).

Kay wondered what elements in Japanese culture could represent a "marquess." She selected "symbol" as the condition in Fig. 1B1, and "element" as the question in Fig. 1B4 (R4). Upon clicking the translation button, [®] the LLM provided two symbols, namely "chrysanthemum" and "family crest." These results aligned with her understanding of Japanese culture. The explanations provided for each symbol facilitated a deeper understanding for Kay. [®] Upon clicking the image generation button, the LLM produced an image aligned with Japanese visual aesthetics (Fig. 1B5) (R5). Consequently, Kay gained insight into the differences and underlying reasons for how "marquess" is represented through elements in Chinese and Japanese cultures (G2).

Culture extrapolation. Kay wanted to know more about symbols representing "marquess" in other countries. In the *target culture extrapolation view* (Fig. 1C), she clicked the inference button to get a series of recommended symbols associated with marquess from different cultures (R6). ⁽¹⁾ For example, the British culture symbolizes a marquess with a coronet mounted with pearls, while Indian and French cultures often associate it with the tiger and the fleur-de-lis.

6.2 Exploring a mysterious TCP

P10 (nicknamed Lily) is an undergraduate student, born and raised in Indonesia, who is strongly interested in Chinese culture but not yet familiar with TCP. This case is highlighted in the supplementary video.

Set target culture information. First, Lily configured the *target culture setting view* to reflect her cultural background (R3), Indonesia. She then evaluated her familiarity with Chinese culture and TCP, assigning ratings of 3 and 1, respectively. After establishing these parameters, she began to explore the system.

Explore composite elements. The lotus is Lily's favorite plant. ① She found the lotus and the egret were closely related in the *element feature view* (R2). She picked the "autumn egret in hibiscus flower landscape" (Fig. 6A1) from the view. ② The other two views in the component provided data distribution of the elements and their symbolic meanings. Furthermore, ③ she discovered that the lotus and egret form a symbolic combination in Chinese culture (Fig. 6A2), representing "auspiciousness" and "nobility," respectively. She continued to learn about the cultural significance of this composite element through the *source culture exploration view*. ④ The combination symbolizes "nobility" because the egret, with its pure white beauty and noble heart, and the lotus, rising in purity from the mud, embody the esteemed quality in Chinese culture (Fig. 6A3) (G1).

Label new elements. The painting's title contains "hibiscus flower," but it was not annotated automatically. (5) Lily examined and eventually identified the hibiscus within the painting (Fig. 6B1) (R1). Then, her interest was directed to its cultural significance (Fig. 6B2). In the *culture exchange component*, she translated the "element" from the source

culture to the "symbol" in the target culture. (C) She discovered through LLMs that in Indonesian culture, the hibiscus flower symbolizes "kecantikan [beauty]" and "kesucian [purity]" (Fig. 6B3). This enhanced her understanding and appreciation of her culture (G2).

"What is a lion dragon?" When catching a glimpse of "lion dragon" in the *word cloud view*, Lily felt extremely curious. "*What is a 'lion dragon?' Both lion and dragon, or neither of them?*" Then, she selected a painting containing the "lion dragon" (Fig. 7A). By carefully observing the painting and conversing for LLM-generated contents in texts and images (Fig. 7D), ⑦ she roughly understood the mythological creature "lion dragon" (G1, R5). ⑧ She noticed that it is associated with the symbol "to ward off evil spirits" (Fig. 7C). ⑨ Inspired by this, she wondered what elements in Indonesia represent the same meaning (Fig. 7E) and successfully got the result: "garuda" and "wayang [shadow puppet]" with their explanations (Fig. 7E1, E2) (R4).

Validate norm acceptability. Her knowledge of these two elements was limited to a few traditional stories she remembered. Consequently, she chose one of the findings and clicked the "appropriate" button in the *target culture extrapolate view* to validate it (R7). The system generated a prompt that included all the information she needed. After submitting it to LLM, **(1)** she received highly positive feedback, completed with detailed reasons and additional recommendations. Drawing on the traditional stories she recalled and the explanations provided by the system, she concluded that this adaptation was acceptable (G3).

Finally, within a 30-minute exploration period, she explored four TCPs and translated two cultural contexts. She was excited that the system offered her a new perspective on TCP and enabled her to discover fascinating insights from other cultures (R6).

6.3 User study with diverse backgrounds

Participants. We recruited ten participants (P1-P10) from eight countries, striving for a balanced distribution across geographical and ethnic origins. Their cultural backgrounds are described in Fig. 9. None of the participants had prior experience with the system. Each participant signed written consent and received a compensation of \$13.80.

Comparative study design. This study adopted a within-subject design with balanced cross-over conditions to counter-balance and control for learning effects. The "*Three Friends of Winter*" [5] and "*Apricot Blossom and Peacock*" [1] were chosen as the subjects. Participants were asked to explore these paintings serendipitously under different conditions and seek cultural insights. They are split into two groups: P1-P5 explored the first painting under the baseline condition and the second with CultiVerse, and vice versa for P6-P10. The order of conditions was randomized.

Baseline condition. The baseline simulates non-natives' typical encounters with TCPs (*e.g.*, museum visits and online media). To parallel the *source culture extraction* component in CultiVerse, participants watched a 2-minute expert commentary on the selected painting. The commentaries were created by an expert who annotated the *TCP-CND* and instructed to explain the paintings to those unfamiliar with Chinese culture. Following the video, participants can access untailored LLMs (same models as CultiVerse without designed prompts) and web search engines to support subsequent exploration.



Fig. 8: The engagement-based questionnaire results regarding the effectiveness and usability of baseline and CultiVerse.



Fig. 9: The number of TCP that participants explored, the number of target culture explored and the number of insights during the openended exploration. The cultural backgrounds of participants are listed.

Procedures. Each study session lasted 90 minutes. It began with a 10-minute introduction to the study background and collection of participants' demographic information and consent. Then, they started the first assigned condition for 10 minutes, followed by a 5-minute feedback session. Participants answered semi-structured interview questions and completed a five-point Likert-scale questionnaire based on the user engagement evaluation theory [36]. The second condition and feedback session took another 15 minutes. CultiVerse was introduced with an example-based introduction [59] before the start of that condition. After experiencing both conditions, participants engaged in a 30-minute open-ended exploration session with CultiVerse. The session finished with a final debriefing to collect comprehensive feedback.

6.4 Results

The assessment for CultiVerse is consistently positive across all measured aspects, as shown in Fig. 8. Detailed findings are outlined below.

Perceived usability assesses whether users can successfully and smoothly perform the desired tasks. All users reported that CultiVerse's analytical flow was clear and straightforward for enhanced exploration efficiency. P9 commented, "*The system compares Chinese culture with the target culture simply and intuitively.*" P2 also remarked on the rapid exploration capabilities, stating, "*I reviewed 15 paintings and their symbolic meanings within 30 minutes, which gave me a good understanding of TCP.*" This efficiency contrasts starkly with the baseline, where participants found it challenging to articulate their inquiries on an unfamiliar topic. The well-designed prompts in CultiVerse evidently enriched the dialogue between users and the system.

Aesthetics evaluates users' impressions of the system's visual and interaction design. Positive feedback, primarily from P1, P3-P6, P8, and P9, concentrated on the *source culture extraction* component. P8 complimented the neat arrangement of paintings and objects, which provided visual comfort. P3 particularly appreciated the *element feature view*'s ability to *"intuitively understand the multi-layered meanings*"

of the composite objects." These designs effectively facilitated the comparison of cultural differences and similarities. However, some feedback indicated that the LLM-generated images lacked authenticity compared to those from search engines. Meeting aesthetic expectations sometimes required multiple image regenerations and prompt engineering, particularly for less-known concepts.

Focused attention assesses users' perception of time passage and concentration during the exploration. All participants have positive ratings on this dimension and have shown enhanced usage efficiency. For instance, P2 highlighted the integration of LLM functionality in the *cultural exchange view*, which facilitated context-based dialogue and helped focus on elaborating specific cultural norms for deeper understanding. Conversely, the need for frequent scrolling back to previous conversations in the baseline condition was seen as a distraction from maintaining focus on the artwork. P10 stated, "*Compared to traditional methods that require frequent switching between web pages, CultiVerse allows me to complete all explorations smoothly.*"

Felt involvement reflects users' evaluation of the overall exploration experience. In the baseline, participants used direct keywords from the commentaries (*e.g.*, painter's name) to guide their searches and dialogues with LLMs, leading to a uniform set of findings. They tended to end exploration once they fully understood the commentary rather than further exploring based on personal interests. In contrast, CultiVerse users reported a heightened sense of engagement due to the opportunity to expand their knowledge boundaries beyond the initial information (P2, P3, and P7-P10). P10 described this cultural exchange process as addictive, similar to "*completing game tasks*," due to the thrill of discovering new insights through experimenting with various combinations. However, P1, P4, and P9 pointed out that selecting different cultural norms in the *cultural exchange* component posed a steep learning curve, impacting their overall experimence.

Novelty evaluates if users can discover surprising information. All participants reported gaining more cross-cultural insights than they did with the baseline. For instance, P7 discovered that crabs symbolize aggressive and overbearing behaviors through their unique sideways movement. These nuanced cultural symbols are often inadequately translated or explained on the Internet. Although LLMs in the baseline provided some support, users struggled with prompt refinements to get the desired results. In contrast, CultiVerse assembles effective tools and customized prompts that facilitate richer cultural exchanges. The explicit option to visualize concepts, such as the tailored representation of crabs moving sideways, has proven effective in helping users understand and remember these cultural symbols.

Endurability refers to the likelihood of users revisiting and recommending the system to others. P10 mentioned considering CultiVerse's information as a reference when choosing gifts for Chinese friends. However, P2 observed that the LLM-generated content had stereotypical impressions, such as associating American culture with eagles and Australian culture closely with kangaroos. This led to doubts about the system's value and intentions for future usage.

6.5 Open-ended exploration

We reviewed users' exploration data and documented three key variables: the number of explored TCPs, selected target cultures, and reported insights during the free exploration (Fig. 9).

Unexpectedly, P4-P6 explored fewer TCPs and cultures and reported fewer insights than others. Although they have closer cultural distances [24] to China than those from European and Asia-Pacific countries (*e.g.*, P1 and P2), their outcomes were not as anticipated. We identified the following reasons from semi-structured interviews. First, the limitations of LLMs in processing information from lesser-dominant cultures led to inaccurate results and long wait times, reducing user trust and enthusiasm with CultiVerse. For example, P4 reported fewer insights since P4 spent more time transitioning between the Han and Tibetan cultures, which was ineffective because LLMs made numerous incomprehensible errors in Tibetan languages. Similar issues were observed with other less commonly represented languages like Pakistani (used by P5), negatively affecting the user experience. While the results can be translated into English, the translation loss and varied English proficiency of the participants may hinder the depth of understanding possible. Second, participants may prefer an in-depth investigation over broader topics, leading to fewer insights in numbers. P6 dedicated significant time to composing poetry in Nepali for a TCP. P5 focused on reading custom information in the TCP-CND dataset. This resulted in fewer explored paintings, cultures, and reported insights.

As P3 belongs to the source culture in CultiVerse, P3 was expected to gain more insights than others. However, P3's skilled use of LLMs also improved the result. Unlike other participants, P3 relied much less on the TCP-CND dataset. Instead, P3 frequently utilized the LLM's generative imaging and prompting to aid cross-cultural understanding. Due to her familiarity with the source culture, more time was spent on the *culture exchange* and *target culture extrapolation* components. P3 noted, "*Exploring whether known elements also hold special meanings in other countries helps me understand cultural diversity.*" Similarly, P9 explored more cultures by translating the same elements into nine different cultures. P9 believed that doing so could facilitate deeper insights into subtle cultural differences.

7 DISCUSSION

We invited four domain experts (E3, E5-E7) with extensive experience in communication studies (E5), TCP education (E7), and cross-cultural translation (E3 and E6) for professional assessments. In this section, we conclude and discuss the lessons learned and the study's limitations.

Enabling new usage scenarios like the Swiss army knife. All four experts unanimously agreed on the significant potential of the project. Their diverse backgrounds contributed to a broad spectrum of suggested applications, further highlighting CultiVerse's versatility across different professional scenarios akin to the Swiss army knife.

E3 praised CultiVerse for its suitability in classroom environments, particularly for its support of interactive and digital teaching methods that cater to various learning styles. The system promotes a pedagogical shift from traditional teaching methods to a more constructivist approach, encouraging active learning and knowledge construction. E3 and E6 also recognized its effectiveness in introductory courses to reduce entry barriers in cross-cultural topics. Nonetheless, they expressed concerns about LLM's ability to self-validate and provide justified information in higher-level academic settings.

E5 appraised CultiVerse as an invaluable tool for overcoming cultural barriers in communication. It can aid art exhibition tour guides deepen their understanding of artworks and enhance their presentation skills. By translating cultural elements and norms to match the audience's background, tour guides can tailor their explanations to be more relevant and engaging. P3's experience supports this, showing that users with background knowledge can utilize CultiVerse more effectively and efficiently. This suggests that professionals such as tour guides and artists could particularly benefit from its features.

E5 and E7 recognized CultiVerse's usability in supporting creativity and research, especially for "cultural comparison" studies. Based on traditional paintings, CultiVerse can provide motivation and stimulate innovative ideas centered around historical and cultural themes. This efficiency in generating ideas reduces the time designers spend brainstorming. It allows them to devote more energy to the creation process, thereby enriching the variety and quality of their designs.

Generalizability and transferability. Our system exemplifies how existing cultural studies frameworks can be adapted to meet the future requirements of intelligent systems focusing on cross-cultural understanding. We have discovered that personal curiosity is crucial in fostering such understanding, and CultiVerse has effectively stimulated this curiosity among users. E3 noted that the CultiVerse workflow was transferable to other art forms, such as Japanese Haiga and Western religious arts. E5 suggested extending this workflow to the film and multimedia industries to decode visual and auditory languages in movies, dramas, and plays. This could enhance audience engagement by analyzing how symbolic elements, like bamboo or the money tree, reflect different character traits in visual narratives. Meanwhile, E6 saw potential in applying the workflow to poetry, exploring how a poem in one culture could be analyzed through the lens of other cultures' paintings and open new interpretative dimensions. Offering users new perspectives. In our user studies, CultiVerse facilitated a richer dialogue with LLMs by eliciting more diverse and insightful user questions than the baseline. The difference between CultiVerse and the baseline, which use the same LLMs, seems to stem from the distinct materials used for source culture extraction. Typically, experts elaborate on artistic elements such as brush techniques and color palettes, but such details may not resonate with all users in a cross-cultural context. In contrast, CultiVerse's focus on cultural elements encourages users to view art through the lens of cultural norms, which helps them connect more meaningfully with diverse cultural values. This approach has increased felt involvement and serendipitous discoveries by exploring various symbolic elements. We recommend that future systems should prioritize concept-driven exploration over fixed narratives to enhance user engagement across broader audiences.

Providing different interpretation spaces. CultiVerse introduces cultural norms as a means to art appreciation. This approach enhances user's level of comprehension in TCPs by encouraging them to consider not just "what" is depicted but also "why" and "how" it is presented. Participants who were well-versed in the source culture reported gaining more insights, suggesting that familiarity with the subject enhances LLM interaction. However, those less acquainted or from different cultural backgrounds do not necessarily gain "fewer" insights but also encounter unique serendipitous findings. For instance, P6 created Nepali poems inspired by traditional Chinese paintings, highlighting a unique cultural exchange. Although introducing cultural norms makes the system's learning curve steeper, this complexity is crucial for facilitating the sophisticated use of LLMs. To assist users, we suggest starting with example-driven explorations to ease them into the system.

Enhancing user accessibility. We noticed that LLMs struggle with cultural awareness (*e.g.*, stereotypes) and bridging language and cultural barriers for less dominant cultures. We have leveraged the innate cultural knowledge of users to reduce these challenges, enabling them to connect with unfamiliar cultures using their prior knowledge. Culti-Verse's streamlined workflow aids users in cross-cultural comparisons and achieving satisfactory outcomes. It offers a one-stop, easy-to-use interface that supports context-based exploration with reference-based prompt engineering [17]. These preset prompts help maintain focus and serve as conversation starters, easing users into interactions with the system, especially those new to LLMs. Moreover, the accurate information from the TCP-CND dataset ensures reliable results even when LLMs underperform. Incorporating multimodal outputs also enhances the novelty and aesthetics of the system.

7.1 Limitations and future work

In our study of cross-cultural understanding through art, several limitations emerged that need addressing. Firstly, our research did not consider participants who have bi-cultural or bilingual backgrounds, whose perspectives could enrich our findings. Additionally, the diversity and quantity of participants were limited, and we assumed that participants had a strong understanding of their own culture, which may not always be the case. Going forward, we plan to use crowdsourcing platforms for broader recruitment in future studies.

E3, E6, and E7 suggested that the current system struggles to capture the deeper symbolic meanings essential for comprehensive cultural understanding. Accurately identifying subtle symbols remains challenging. For instance, a hooded falcon represents the opposite meaning of a flying one, but identifying a hooded falcon is extremely difficult. This requires more sophisticated model training in recognizing such elements. Moreover, improving the system's ability to process additional contextual information like historical, geographical, and textual details on artworks is promising for a richer semantic and multimodal interpretation. Our approach to validating cultural norms using sentiment-based emotions needs improvement, especially in handling composite elements. E6 also emphasized the system's shortcomings in aligning historical and cultural contexts, with LLMs often misinterpreting historical contexts through contemporary cultural perspectives. This mismatch and the concerns noted by P2 about stereotypes are crucial areas for future development.

FIGURE CREDITS

Fig. 1B3, Fig. 1B5, Fig. 7D, Fig. 7E1, and Fig. 7E2 are generated by OpenAI's GPT4-Turbo.

REFERENCES

- [1] Apricot Blossom and Peacock. https://digitalarchive.npm.gov. tw/Painting/Content?pid=3596&Dept=P. 7
- [2] New models and developer products announced at DevDay. https://openai.com/blog/new-models-and-developer-productsannounced-at-devday. [Accessed Dec 1, 2023]. 3
- [3] The Chinese Iconography Thesaurus Database. https: //chineseiconography.org. [Accessed Sep 1, 2023]. 4
- [4] The GitHub URL for the CultiVerse project. https://github.com/ cultiverse/CultiVerse. 2
- [5] The Three Friends of Winter. https://digitalarchive.npm.gov. tw/Painting/Content?pid=5872&Dept=P. 7
- [6] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat, et al. Gpt-4 technical report. arXiv preprint arXiv:2303.08774, 2023. doi: 10.48550/arXiv.2303 .08774 3
- [7] M. F. Adilazuarda, S. Mukherjee, P. Lavania, S. Singh, A. Dwivedi, A. F. Aji, J. O'Neill, A. Modi, and M. Choudhury. Towards measuring and modeling" culture" in llms: A survey. *arXiv preprint arXiv:2403.15412*, 2024. doi: 10.48550/arXiv.2403.15412 3
- [8] J. J. Amigo Extremera. Fitting culture into translation process research. *Translation & Interpreting: The International Journal of Translation and Interpreting Research*, 7(1):26–46, 2015. doi: 10.12807/ti.106201.2015. a05 3
- [9] S. Bassnett. Culture and translation. A companion to translation studies, pp. 13–23, 2007. doi: 10.21832/9781853599583-003 3
- [10] G. Beech. Misunderstanding misunderstanding: Exploring the depth of hindrance to cross-cultural communications. *International Bulletin of Mission Research*, 47:274–285, 2022. doi: 10.1177/23969393221105220 3
- [11] A. J. Bradley, M. El-Assady, K. Coles, E. Alexander, M. Chen, C. Collins, S. JÃd'nicke, and D. J. Wrisley. Visualization and the digital humanities:. *IEEE Computer Graphics and Applications*, 38(6):26–38, 2018. doi: 10. 1109/MCG.2018.2878900 2
- [12] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. doi: 10.48550/arXiv.2005.14165 3
- [13] S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, et al. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*, 2023. doi: 10.48550/arXiv.2303.12712 3
- [14] X. Chen, Y. Chen, T. Tang, X. Ge, R. Wang, Y. Wang, and X. Wang. A multi-modal visual analytic approach for attributing and authenticating ancient chinese paintings. *Journal of Computer-Aided Design and Computer Graphics*, p. to appear, 2023. doi: 10.3724/SP.J.1089.2023-00350 2
- [15] W. Eberhard. Chinese Cultural Symbol Dictionary. Hunan Literature and Art Publishing House, 1990. 4
- [16] Y. Feng, J. Chen, K. Huang, W. Kam-Kwai, H. Ye, W. Zhang, R. Zhu, X. Luo, and W. Chen. ipoet: Interactive painting poetry creation with visual multimodal analysis. *Journal of Visualization*, 25(3):671–685, June 2022. doi: 10.1007/s12650-021-00780-0 1
- [17] Y. Feng, X. Wang, W. Kam-Kwai, S. Wang, Y. Lu, M. Zhu, B. Wang, and W. Chen. Promptmagician: Interactive prompt engineering for textto-image creation. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):295–305, 2024. doi: 10.1109/TVCG.2023.3327168 9
- [18] Y. Feng, X. Wang, B. Pan, W. Kam-Kwai, Y. Ren, S. Liu, Z. Yan, Y. Ma, H. Qu, and W. Chen. Xnli: Explaining and diagnosing nli-based visual data analysis. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–14, 2023. doi: 10.1109/TVCG.2023.3240003 3
- [19] W. C. Fong. How to understand chinese painting. Proceedings of the American Philosophical Society, 115(4):282–292, 1971. 3
- [20] Y. R. Fung, T. Chakrabarty, H. Guo, O. Rambow, S. Muresan, and H. Ji. NORMSAGE: Multi-lingual multi-cultural norm discovery from conversations on-the-fly. In *Proc. EMNLP*, EMNLP'23, pp. 15217–15230, Dec. 2023. doi: 10.18653/v1/2023.emnlp-main.941 3
- [21] Y. Guo, Y. Luo, K. Lu, L. Li, H. Yang, and X. Yuan. Liberroad: Probing into the journey of chinese classics through visual analytics. *IEEE Trans-*

actions on Visualization and Computer Graphics, 30(1):529–539, 2024. doi: 10.1109/TVCG.2023.3326944 2

- [22] R. J. Hallman. The art object in hindu aesthetics. *The Journal of Aesthetics and Art Criticism*, 12(4):493–498, 1954. doi: 10.2307/426909 2
- [23] A. Harris. Archival objects and the circulation of culture. *Circulating Cultures*, 2014. doi: 10.22459/CC.12.2014.01 2
- [24] G. Hofstede. Dimensionalizing cultures: The hofstede model in context. Online readings in psychology and culture, 2(1):8, 2011. doi: 10.9707/ 2307-0919.1014 8
- [25] J. House. Misunderstanding in intercultural communication. *Toegepaste Taalwetenschap in Artikelen*, 57:11–17, 1997. doi: 10.1075/TTWIA.57. 02HOU 3
- [26] J. Huang and D. Yang. Culturally aware natural language inference. In Findings of the Association for Computational Linguistics: EMNLP 2023, pp. 7591–7609, 2023. doi: 10.18653/v1/2023.findings-emnlp.509 3
- [27] W. Kam-Kwai, Y. Luo, X. Yue, W. Chen, and H. Qu. Prismatic: Interactive multi-view cluster analysis of concept stocks. arXiv preprint arXiv:2304.05011, 2024. doi: 10.48550/arXiv.2402.08978 3
- [28] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei. Scaling laws for neural language models. arXiv preprint arXiv:2001.08361, 2020. doi: 10.48550/ arXiv.2001.08361 3
- [29] M. J. Letts. Cross-cultural issues in art: Frames for understanding. *Studies in Art Education*, 56(2):187–190, 2015. doi: 10.1080/00393541.2015. 11518961 2
- [30] S. Liu et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. arXiv preprint arXiv:2303.05499, 2023. doi: 10.48550/arXiv.2303.05499 4
- [31] Y. Liu. Cross-culturalism in painting: visualization via meanders. Journal of Visual Art Practice, 8:205–214, 2009. doi: 10.1386/jvap.8.3.205/1 2
- [32] Z. Liu, Y. Shin, B.-C. Okogwu, Y. Yun, L. Coleman, P. Schaldenbrand, J. Kim, and J. Oh. Towards equitable representation in text-to-image synthesis models with the cross-cultural understanding benchmark (ccub) dataset. arXiv preprint arXiv:2301.12073, 2023. doi: 10.48550/arXiv. 2301.12073 3
- [33] E. Martinez. Listen to that painting!: An intercultural experiment based on cross-sensory coupling. *The International Journal of New Media*, *Technology and the Arts*, 7:33–47, 2013. doi: 10.18848/2326-9987/CGP/ V07103/36307 2
- [34] N. P. Museum. National Palace Museum Open Data. https://theme. npm.edu.tw/opendata/. [Accessed Sep 1, 2023]. 4
- [35] T. Naous, M. J. Ryan, and W. Xu. Having beer after prayer? measuring cultural bias in large language models. *arXiv preprint arXiv:2305.14456*, 2023. doi: 10.48550/arXiv.2305.14456 3
- [36] H. L. O'Brien and E. G. Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1):50–69, 2010. doi: 10.1002/asi. 21229 8
- [37] Y. Pan and W. Yu. Discussion on aesthetic design in chinese painting based on cross-cultural design. In *Proc. HCII*, Poster HCII 2020, pp. 308–316, 2020. doi: 10.1007/978-3-030-50732-9_41 3
- [38] S. Rai, K. Z. Zaveri, S. Havaldar, S. Nema, L. Ungar, and S. C. Guntuku. A cross-cultural analysis of social norms in bollywood and hollywood movies. arXiv preprint arXiv:2402.11333, 2024. doi: 10.48550/arXiv. 2402.11333 3
- [39] A. Ramezani and Y. Xu. Knowledge of cultural moral norms in large language models. arXiv preprint arXiv:2306.01857, 2023. doi: 10.18653/ v1/2023.acl-long.26 3
- [40] I. I. T. Rohmah. Cross Cultural Understanding (The Road To Travel The World). Perkumpulan Rumah Cemerlang Indonesia, 2021. 2
- [41] K. Roth. Material culture and intercultural communication. International Journal of Intercultural Relations, 25:563–580, 2001. doi: 10.1016/S0147 -1767(01)00023-2 2
- [42] Y. Ruan. Exploring the symbolism and cultural importance of dunhuang lotus motifs. *International Journal of Education and Humanities*, 10(3):22– 24, 2023. doi: 10.54097/ijeh.v10i3.11785 2
- [43] E.-M. Sent and A. L. Kroese. Commemorating geert hofstede, a pioneer in the study of culture and institutions. *Journal of Institutional Economics*, 18(1):15–27, 2022. 4
- [44] F. G. Sharipboevna and T. Kamola. Improving cross-cultural communication. *ResearchJet Journal of Analysis and Inventions*, 2:359–363, 2021. doi: 10.17605/OSF.IO/9KWDM 2
- [45] C.-W. Sky, A. Saakyan, O. Li, Z. Yu, and S. Muresan. Sociocultural

norm similarities and differences via situational alignment and explainable textual entailment. In *Proc. EMNLP*, 2023. doi: 10.48550/arXiv.2305. 14492 3

- [46] H. Sun. Creation and interpretation: Study on the symbolism of traditional chinese painting. In *Proc. ICASSEE*, vol. 1 of *ICASSEE*'21, pp. 240– 246. Amsterdam University Press, 2021. doi: 10.5117/9789048557240/ ICASSEE.2021.035 3
- [47] J. Sun and A. Ab Aziz. Exploring the redesign value of chinese traditional auspicious patterns in the context of modern design. *EDUCATUM Journal* of Social Sciences, 9(1):12–26, 2023. doi: 10.37134/ejoss.vol9.1.2.2023
- [48] T. Tang, Y. Wu, P. Xia, W. Wu, X. Wang, and Y. Wu. Pcolorizor: Recoloring ancient chinese paintings with ideorealm-congruent poems. In *Proc. UIST*, UIST '23, 2023. doi: 10.1145/3586183.3606814 4
- [49] T. Taracenko and L. Kulykova. The ways of overcoming misunderstanding in cross-cultural communication. *Scientific Bulletin of Bogdan Chmelnitskiy Melitopol State Pedagogical University. Series: Pedagogy*, 2:40–44, 2018. doi: 10.33842/2219-5203-2018-2-21-40-44 3
- [50] J. H. Taylor. Journey through the afterlife: Ancient Egyptian Book of the Dead. Harvard University Press, 2010. 2
- [51] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023. doi: 10.48550/arXiv.2307.09288 3
- [52] J. Von Oswald, E. Niklasson, E. Randazzo, J. Sacramento, A. Mordvintsev, A. Zhmoginov, and M. Vladymyrov. Transformers learn in-context by gradient descent. In *International Conference on Machine Learning*, pp. 35151–35174. PMLR, 2023. doi: 10.48550/arXiv.2212.07677 3
- [53] Y. Wang, H. Liang, X. Shu, J. Wang, K. Xu, Z. Deng, C. Campbell, B. Chen, Y. Wu, and H. Qu. Interactive visual exploration of longitudinal historical career mobility data. *IEEE Transactions on Visualization and Computer Graphics*, 28(10):3441–3455, 2021. doi: 10.1109/TVCG.2021. 3067200 1, 6
- [54] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022. doi: 10. 48550/arXiv.2206.07682 3
- [55] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022. doi: 10.48550/arXiv.2201.11903 3
- [56] C. A. S. Williams. Outlines of Chinese symbolism and art motives: an alphabetical compendium of antique legends and beliefs, as reflected in the manners and customs of the Chinese. Courier Corporation, 1976. 3
- [57] C. A. S. Williams. Chinese symbolism and art motifs fourth revised edition: A comprehensive handbook on symbolism in Chinese art through the ages. Tuttle Publishing, 2012. 3
- [58] R. Xu, Y. Sun, M. Ren, S. Guo, R. Pan, H. Lin, L. Sun, and X. Han. Ai for social science and social science of ai: A survey. *Information Processing* & *Management*, 61(3):103665, 2024. doi: 10.1016/j.ipm.2024.103665
- [59] L. Yang, C. Xiong, W. Kam-Kwai, A. Wu, and H. Qu. Explaining with examples lessons learned from crowdsourced introductory description of information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(3):1638–1650, 2023. doi: 10.1109/TVCG.2021. 3128157 8
- [60] S. Ye, C. Zhu-Tian, X. Chu, K. Li, J. Luo, Y. Li, G. Geng, and Y. Wu. Puzzlefixer: A visual reassembly system for immersive fragments restoration. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):429– 439, 2023. doi: 10.1109/TVCG.2022.3209388 2
- [61] L. Yuan. Cultural differences of chinese loong and western dragon. Studies in Literature and Language, 10(3):40, 2015. 2
- [62] S. Zhang, L. Dong, X. Li, S. Zhang, X. Sun, S. Wang, J. Li, R. Hu, T. Zhang, F. Wu, et al. Instruction tuning for large language models: A survey. arXiv preprint arXiv:2308.10792, 2023. doi: 10.48550/arXiv. 2308.10792 3
- [63] W. Zhang, W. Kam-Kwai, Y. Chen, A. Jia, L. Wang, J.-W. Zhang, L. Cheng, H. Qu, and W. Chen. Scrolltimes: Tracing the provenance of paintings as a window into history. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–11, 2024. doi: 10.1109/TVCG.2024.3388523 1, 2
- [64] W. Zhang, W. Kam-Kwai, X. Wang, Y. Gong, R. Zhu, K. Liu, Z. Yan, S. Tan, H. Qu, S. Chen, and W. Chen. Cohortva: A visual analytic system for interactive exploration of cohorts based on historical data. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):756–766,

2023. doi: 10.1109/TVCG.2022.3209483 1, 2

- [65] W. Zhang, Q. Ma, R. Pan, and W. Chen. Visual storytelling of song ci and the poets in the social–cultural context of song dynasty. *Visual Informatics*, 5(4):34–40, 2021. doi: 10.1016/j.visinf.2021.12.002 1, 2
- [66] W. Zhang, S. Tan, S. Chen, L. Meng, T. Zhang, R. Zhu, and W. Chen. Visual reasoning for uncertainty in spatio-temporal events of historical figures. *IEEE Transactions on Visualization and Computer Graphics*, 29(6):3009–3023, 2023. doi: 10.1109/TVCG.2022.3146508 1
- [67] W. Zhang, J. Zhang, W. Kam-Kwai, Y. Wang, Y. Feng, L. Wang, and W. Chen. Computational approaches for traditional chinese painting: From the "six principles of painting" perspective. *Journal of Computer Science and Technology*, 2024. To appear. doi: 10.48550/arXiv.2307. 14227 3
- [68] Y. Zhang, Y. Li, L. Cui, D. Cai, L. Liu, T. Fu, X. Huang, E. Zhao, Y. Zhang, Y. Chen, et al. Siren's song in the ai ocean: a survey on hallucination in large language models. *arXiv preprint arXiv:2309.01219*, 2023. doi: 10. 48550/arXiv.2309.01219 4
- [69] S. Zhao, D. Kirk, S. Bowen, and P. Wright. Enhancing the appreciation of traditional chinese painting using interactive technology. *Multimodal Technologies and Interaction*, 2(2):2414–4088, 2018. doi: 10. 3390/mti2020016 3
- [70] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, et al. A survey of large language models. *arXiv preprint* arXiv:2303.18223, 2023. doi: 10.48550/arXiv.2303.18223 3