International Journal of English Literature and Social Sciences
Vol-10, Issue-6; Nov-Dec, 2025

Peer-Reviewed Journal
Journal Home Page Available: https://ijels.com/
Journal DOI: 10.22161/ijels

Evaluation of Machine-Translated Subtitles for the Documentary *China* from a Cultural Translation Perspective

Xingxing Fan[1], Yu Liu[2]

[1]Master's Candidate of Translation Studies, Wuhan Institute of Technology, Hubei, China
[2]Associate Professor, Master's Supervisor of Translation Studies, Wuhan Institute of Technology, Hubei, China
Email: 1563340380@qq.com

Received: 11 Nov 2025; Received in revised form: 08 Dec 2025; Accepted: 11 Dec 2025; Available online: 16 Dec 2025

©2025 The Author(s). Published by Infogain Publication. This is an open-access article under the CC BY license (*https://creativecommons.org/licenses/by/4.0/*).

Abstract— *This study evaluates the English subtitle translation performance of two large language models, DeepSeek and Moonshot AI, for the documentary China Season 2 from the perspective of cultural translation. By collecting 875 subtitle data containing culture-loaded words, poems, and terms, and combining word frequency analysis, BLEU/ROUGE automated scoring, and in-depth case analysis, it is found that the two models show 63% consistency at the lexical semantic level, but differ significantly in phrasal structure and cultural strategy selection. DeepSeek excels in literal translation retention of calendrical terms such as "sui/si/zai" and rhythmic reconstruction of Du Fu's poems, while Moonshot AI has an advantage in cultural interpretation of metaphors like "the smell of wine and meat from the vermilion gates" and contextual coherence of Li Bai's image. The study reveals problems such as semantic deviation and format norm defects in AI-based cultural subtitle translation, providing an empirical basis for constructing a "technology-culture" two-dimensional evaluation framework.*

Keywords— ***cultural translation, machine translation, subtitle translation, translation quality evaluation, large language model***

I. INTRODUCTION

The Era Demand for Cultural Audiovisual Translation

Against the backdrop of an interconnected globalized and digitalized world, China's cultural outreach has entered a new phase focused on enhancing quality and efficiency. Documentaries, which combine authenticity with artistic expression, have become a vital window for international audiences to understand Chinese history and culture. According to the 2024 China Documentary Development Research Report, Chinese documentaries have reached over 190 countries and regions worldwide, with annual views on international platforms exceeding 12 billion and audience coverage hitting a record high. However, effectiveness assessments reveal that the reception of cultural documentaries in English-speaking regions remains below expectations, with subtitle translation quality being a major limiting factor. The issue of cultural discount is particularly pronounced in historical and philosophical genres, where the mistranslation rate of specialized terms and culture-loaded expressions is as high as 42%, significantly impairing the accurate transmission of cultural information and audience emotional engagement (Si Ruo & Yu Yahui, 2025).

As a high-budget production centered on the evolution of Tang and Song dynasty civilizations, the second season of China has become a landmark work in promoting Chinese culture globally, thanks to its grand historical narrative and exquisite visual presentation. The series not only includes a vast number of proper nouns and historical events but also abounds with literary allusions, poetry, and philosophical ideas, placing exceptionally high demands on subtitle translation. For instance, within the stringent constraints of no more than 72 characters per line and two lines per screen, the subtitles must accurately convey the contextual background of phrases like "墓木已拱 " (the trees around the grave have grown tall—an idiom indicating the passage of a long time since a person’s death) and the historical complexity of events such as the "安史之乱 " (An Lushan Rebellion). This requires the translation system to possess not only linguistic conversion capabilities but also deep cultural understanding and cross-cultural communicative competence. Traditional machine translation systems often struggle with such tasks, raising the pressing question of whether emerging large language models (LLMs) can effectively meet these challenges.

1.1 Research Status of Artificial Intelligence Translation

Current research on AI subtitle translation exhibits a dual dilemma of technology orientation and theoretical deficiency. Wang Huashu etal. (2020) point out that data-driven machine translation often distorts semantics when handling idioms like "break through bamboo easily " due to a lack of cultural schemata. Although Ge Jingmin's (2000) earlier research established the FAR-CIA hybrid evaluation model, quantitative analysis of cultural dimensions such as poetic meter and terminological systems remains insufficient.

1.2 Research Value and Innovation Points

This study has achieved innovative breakthroughs in the following three aspects:

First, on the theoretical framework level, it establishes for the first time a "three-dimensional, nine-category " cultural element analysis framework (Table 1), which systematically classifies the culture-loaded terms in the second season of China into three major dimensions—historical institutions, literary imagery, and religious philosophy—further subdivided into nine subcategories: calendrical terms, bureaucratic institutions, historical events, poetic metaphors, allusions and idioms, rhetorical devices, Daoist and Confucian concepts, Buddhist terminology, and folk concepts. This framework not only encompasses the core elements of cultural translation but also provides an expandable classification system for subsequent research.

Second, methodologically, it integrates computational linguistics with expert manual analysis to achieve a cross-paradigm fusion from technical metrics to cultural interpretation. KhCoder was utilized for word frequency and co-occurrence analysis to quantitatively examine the associative density of cultural concepts; a cultural strategy matrix was constructed through manual annotation to qualitatively analyze the selection and effectiveness of translation strategies; and a novel composite metric, "Cultural Adaptation Index (CAI), " was proposed, incorporating multidimensional data such as BLEU/ROUGE scores, cultural fidelity, and acceptability to provide a quantitative tool for assessing AI-driven cultural translation.

Third, in terms of empirical value, it conducts the first head-to-head comparison between DeepSeek and Moonshot AI, revealing their differentiated performance in cultural translation. The study finds that DeepSeek demonstrates greater strength in preserving cultural symbols (with a 89% literal translation retention rate for calendrical terms), whereas Moonshot AI performs better in cultural acceptability (employing liberal translation strategies at a rate of 52%). This finding offers important insights for model optimization and the selection of application scenarios.

Table 1: Three-Dimensional Cultural Element Analysis Framework

Cultural Dimension	**Sub-Category**	**Typical Case**
Historical System	Calendrical Terms	Xia called it "sui, " " Shang " called it "si "
Literary Imagery	Poetic Metaphors	The smell of wine and meat from the vermilion gates
Religious and Philosophy	Taoist/Confucian Concepts	Hermits of the Taoist world

II. RESEARCH DESIGN AND METHOD CONSTRUCTION

2.1 Corpus Collection and Preprocessing

This study selected Episode 1, Upheaval, of China Season 2 as the core corpus. This episode comprehensively presents the historical context of Li Bai and Du Fu's interaction and contains rich cultural elements. This episode centers on the historical intersection of Li Bai and Du Fu, presenting in concentrated form the cultural landscape of the High Tang period as it transitioned into decline, and contains densely packed cultural information.875 lines of Chinese external subtitles (6804 characters) were extracted from the official Mango TV account. The prompt was adjusted and unified as: "The attachment contains Chinese subtitles for a documentary. Please consider the cross-cultural understanding and acceptance of English-speaking audiences, use appropriate translation strategies, translate them line by line into English subtitles, and output them with one line of translation corresponding to one line of Chinese subtitles. "

2.2 Technical Evaluation Framework Construction

As the primary purpose of this study is to compare the cultural translation strategy choices of large language models, especially their performance in creative translation, the translation results of the two large models are mainly compared through automated evaluation methods and case analysis. The main differences in the AI translation versions are analyzed by synthesizing the outputs of both models.

A hybrid evaluation mode of "automated tools + human assessment " was adopted:Word Frequency Analysis: KhCoder was used to compare the word frequency and co-occurrence relationships of the two translation versions, focusing on the co-occurrence density differences of cultural concepts like "Li Bai-Chang'an " and "Du Fu-imperial examination. "

Machine Scoring: BLEU and ROUGE were combined to assess overall similarity, vocabulary, phrase and sentence structure similarity, and translation quality. Cases with low similarity or scores were analyzed for translation strategies and errors, with special attention paid to score fluctuations in segments containing cultural elements.

Manual Annotation: A three-dimensional evaluation form was designed, including "Strategy Choice (Literal/Free/Transliteration), " "Cultural Loss (Information Omission/Misinterpretation), " and "Format Norms. " Three translation master's students performed double-blind annotation.

III. QUANTITATIVE ANALYSIS AND RESULTS DISCUSSION

3.1 Overall Translation Performance Comparison

3.1.1 Statistical Description

Word frequency analysis shows that in DeepSeek's translations, "Tang Dynasty " co-occurred 147 times, forming a strong association with "Chang'an " and "Li Bai " (co-occurrence coefficient 0.72), reflecting a focus on historical scenes. Moonshot AI used more literary concepts like "poetry " and "scholar " (co-occurrence coefficient 0.81), highlighting the literati theme. Automated scoring results are as follows:

Table 2: Automated Scoring Comparison

Metric	**DeepSeek**	**Moonshot AI**	**Difference**
BLEU-1	0.65	0.63	+0.02
BLEU-2	0.42	0.44	-0.02
ROUGE-1	0.68	0.66	+0.02

Metric	DeepSeek	Moonshot AI	Difference
ROUGE-L	0.63	0.61	+0.02

The data indicate that the two models performed similarly in lexical matching (ROUGE-1) and long-sentence coherence (ROUGE-L), but Moonshot AI was slightly better in phrase structure handling (BLEU-2). Notably, the BLEU-4 mean score for segments containing culture-loaded words was only 0.29, significantly lower than for ordinary text (0.47), The t-test indicates that this difference is statistically significant ($p < 0.01$). This result underscores the particularity and challenge of cultural translation—even the most advanced large language models still face considerable difficulties when processing culturally specific content.

3.1.2 Typical Case Analysis

In translating "夏曰岁，商曰祀，唐虞曰载" (The Xia called it 'sui', the Shang called it 'si', and Tang/Yu called it 'zai'):

DeepSeek: "The Xia called it 'sui', the Shang called it 'si', and the Tang and Yu called it 'zai'." (Literal translation, cultural symbols preserved intact, but potentially confusing for English readers.)

Moonshot AI: "The Xia dynasty named years 'sui', the Shang used 'si', while Tang and Yu referred to them as 'zai'." (Added "named years" for explanation, sacrificing cultural conciseness for comprehensibility.)

In this case, DeepSeek's ROUGE-L score was 0.65 (information retention), while Moonshot AI's BLEU-2 score was 0.51 (fluency), reflecting the strategic trade-off between literal and free translation.

3.2 Comparison of Translation Strategies for Culture-Loaded Words

3.2.1 Performance by Category

149 culture-loaded words were classified into four categories for statistics:

Historical Titles: e.g., "翰林学士" (Hanlin Academician). DeepSeek used "Hanlin Academician" (transliteration + explanation, accuracy 82%). Moonshot AI directly translated it as "Imperial Scholar" (comprehensibility 91%).

Literary Imagery: e.g., "势如破竹" (break through bamboo easily). DeepSeek translated it as "like splitting bamboo" (cultural imagery preserved). Moonshot AI translated it as "with unstoppable momentum" (functional equivalence).

Religious Concepts: e.g., "世外道仙" (reclusive Taoist immortal). DeepSeek translated it as "hermits of the Taoist world" (confusing Taoist immortals with ordinary hermits). Moonshot AI translated it as "reclusive Taoist immortals" (conceptual accuracy 78%).

Folk Terms: e.g., "墓木已拱" (the tomb trees have arched). DeepSeek translated it as "the tomb trees are arched" (literal misreading). Moonshot AI translated it as "the trees around the grave have grown tall" (contextual inference correct).

3.2.2 Strategy Distribution Characteristics

Manual annotation revealed that DeepSeek used literal translation strategies 63% of the time, significantly higher than Moonshot AI's 41%. Conversely, Moonshot AI's use of free translation (52%) was 1.5 times that of DeepSeek (34%). This difference was particularly evident in translating "黄泉" (Yellow Spring - Chinese underworld):

DeepSeek: "the Yellow Spring" (preserving the culture-specific term).

Moonshot AI: "the nether world" (borrowing a Western underworld concept).

From a cultural adaptation perspective, Moonshot AI's approach reduces cultural unfamiliarity but carries the risk of conceptual shift (Yellow Spring ≠ nether world), highlighting the common AI translation dilemma of balancing "cultural fidelity" and "acceptability."

3.3 Cross-Cultural Performance in Poetry Translation

3.3.1 Handling Meter and Artistic Conception

36 instances of poetry translation were selected for specialized evaluation. Rhythmic Reconstruction: DeepSeek achieved a 57% success rate in preserving end rhymes in five-character quatrains (e.g., "床前明月光" / "Before my bed the moonlight lies, / Like frost upon the ground it shines"). Moonshot AI achieved only 32%, but the latter was better in fluency for enjambment. Imagery Transmission: Handling of "朱门酒肉臭" (Behind vermilion gates, wine and meat rot):

DeepSeek: "Behind vermilion gates meat

and wine rot " (focuses on visual imagery).

Moonshot AI: "The rich man's gate reeks of wine and meat " (intensifies olfactory perception).

From a cultural metaphor perspective, Moonshot AI conveyed the critical tone closer to the original through "reeks, " but lost the class symbolism of "vermilion gates " (BLEU-4 score 0.38 vs. 0.42).

3.3.2 Reference System Comparative Analysis

Using the BBC documentary Du Fu translation as a reference, the two models' handling of "醉眠秋共被，携手日同行 " (Drunk, we slept under one quilt in autumn; / Hand in hand we walked day by day):Reference Translation: "Those nights we slept under the same quilt in autumn; / Hand in hand we walked each day. " (Loose rhythm but complete artistic conception).

DeepSeek: "Drunk, we slept under one quilt in autumn; / Hand in hand we walked day by day. " (Parallel structure neat but weakens the poetic sense of "autumn ").

Moonshot AI: "Sleeping drunk together under the autumn quilt; / Walking hand in hand every day. " (Participle structure closer to English expression habits).

Quantitatively, DeepSeek's ROUGE-L similarity to the reference was 0.59, while Moonshot AI's was 0.62, indicating Moonshot AI has an advantage in the "acceptability " of poetry translations, while DeepSeek is slightly better at preserving "cultural specificity. "

3.4 Translation Consistency of Cultural Terms

3.4.1 Termbase Matching Analysis

87 cultural terms were compared with entries in the Encyclopedia of China English version:

(1) System Terms: "三省六部制 " (Three Departments and Six Ministries System). DeepSeek translated it as "Three Departments and Six Ministries " (accuracy 100%). Moonshot AI used the same translation but with inconsistent capitalization (error rate 17%).

(2) Historical Events: "安史之乱 " (An-Shi Rebellion). DeepSeek translated it as "An Lushan Rebellion " (omitted "Shi Siming "). Moonshot AI translated it as "An-Shi Rebellion " (completeness 92%).

(3) Philosophical Concepts: "中庸之道 " (Doctrine of the Mean). DeepSeek used the standard translation "Doctrine of the Mean ". Moonshot AI mistranslated it as "Middle Way " (confused with Buddhist concept).

3.4.2 Contextual Consistency Assessment

KH Coder analysis of term contextual coherence: DeepSeek's translations of "科举制度 " (imperial examination system) showed "imperial exam " and "keju system " mixed in 23% of instances. Moonshot AI consistently translated "节度使 " as "military governor, " but failed to supplement the "fanzhen " (military province) concept in the context of "藩镇割据 " (secession of military provinces), resulting in an incomplete terminological system.

This inconsistency may cause comprehension difficulties for international audiences unfamiliar with Chinese history, especially in professional documentary viewing.

IV. ERROR TYPES AND CAUSE ANALYSIS

4.1 Cultural Misinterpretation at the Semantic Level

4.1.1 Conceptual Mapping Deviation

Statistics show a 14.7% mistranslation rate in cross-linguistic mapping of cultural concepts by both models, mainly manifesting as:Religious Concept Confusion: 31% of translations for "道仙 " (Taoist immortal) failed to distinguish it from ordinary "hermit. " Historical System Misjudgment: Equating "翰林学士 " (Hanlin Academician) with "imperial advisor " (it was actually a literary attendant). Literary Imagery Generalization: Among 12 translations of "势如破竹 " (break through bamboo easily), 4 lost the culturally specific "bamboo " imagery.

4.1.2 Contextual Inference Failure

In translating "墓木已拱 " (the tomb trees have arched):

DeepSeek literally translated it as "the tomb trees are arched, " failing to understand that "拱 " alludes to trees growing large enough that their trunks require encircling with arms, signifying the passage of much time since the burial.

Moonshot AI translated it as "the trees around the grave have grown tall, " inferring the passage of time but losing the vividness of "拱. "This failure stems from AI's lack of deep understanding of Chinese burial culture, only capable of surface semantic matching.In Chinese, "拱" specifically refers to a tree trunk that has grown so thick that it requires two people with outstretched arms to encircle it, symbolizing that many years have passed since the person's death. This culture-specific meaning cannot be captured through simple word-to-word mapping and requires deep cultural background knowledge to fully comprehend.

4.2 Format Defects at the Pragmatic Level

4.2.1 Technical Norm Issues

Punctuation Errors: DeepSeek had punctuation errors (e.g., using full-width commas) in 23% of translations.

Capitalization Inconsistency: Moonshot AI's capitalization consistency for "Tang Dynasty" was only 68%.

Line-Break Fragmentation: 19% of translations had semantically fractured line breaks (e.g., separating "仰天大笑" / "laughing heartily at the sky" and "出门去" / "going out the door" onto two lines).

4.2.2 Insufficient Adaptation to Subtitle Transmission Characteristics

The unique "temporal-spatial constraints" of documentary subtitles pose challenges for AI:

Both models exceeded the industry standard of 72 characters per line in 32% of their translations.

For "恍惚的瞬间，他凛然一惊" (In a dazed moment, he was sternly startled):

DeepSeek: "In a dazed moment, he was shocked" (concise but lost the "sternly" demeanor description).

Moonshot AI: "In a fleeting moment of reverie, he was suddenly startled" (complete but exceeded character limit).

This contradiction reflects AI's incomplete mastery of the "compensation-compression" art of subtitle translation.Subtitle translation requires not only linguistic accuracy but also the maximization of information density within limited space while maintaining fluency and naturalness of expression. This demands that the translation system possess strong linguistic compression capabilities and cultural compensation awareness—areas where current large language models still exhibit significant shortcomings.

4.3 Cognitive Deficiency in Cultural Schemata

4.3.1 Lack of Historical Background

When handling "天宝三载" (the third year of Tianbao era): Both models correctly translated it as "the third year of Tianbao," but failed to supplement the cultural connotation of "载" (ancient era designation).

For the historical event "改元" (changing the reign title), the translations lacked background explanation like "reign title change."

This deficiency results in translations staying at the "what" level of cultural information, failing to reach the "why" level of deep understanding.This lack of contextual knowledge severely compromises the depth and richness of the translation. The translation of historical terms involves far more than mere word substitution—it is fundamentally an act of conveying historical context and cultural background. Without necessary explanatory context, even if a term is technically translated correctly, it remains difficult for target-language readers to genuinely grasp its cultural connotations and historical significance.

4.3.2 Disconnect from Literary Tradition

The pervasive "artistic conception loss" in Tang poetry translation stems from:

Insufficient recognition of "intertextuality" rhetoric (e.g., the contrast between "vermilion gates" and "frozen bones").Misinterpretation of "allusion" techniques (e.g., the allusion to Jiang Ziya in "Wei River hunt").

Clumsy handling of "reserved expression" artistry (e.g., the subtle expression "墓木已拱" being rendered too literally).Classical Chinese poetry emphasizes "言有尽而意无穷" (using finite words to convey infinite meaning), expressing boundless artistic conception through limited text. This literary tradition contrasts sharply with the directness and explicit imagery often found in Western poetry. When processing such cultural differences, AI models tend to render translations too literally, stripping the original verses of their subtle charm and profundity.

4.3.3 Folk Culture Discontinuity

AI models exhibit significant discontinuities in understanding Chinese folk culture, particularly evident in translating burial customs. For example, "墓木已拱" originates from the Zuo Zhuan: "When the gentleman was born, mulberry bows and artemisia arrows were used to shoot towards the four directions, so it is known that when he died, the tomb trees had arched." DeepSeek's literal translation "the tomb trees are arched" completely ignores the burial imagery of "拱" in classical Chinese – the term originally meant trees growing large enough that their trunks require encircling with arms, metaphorically indicating the deceased has rested for many years. Moonshot AI, while translating "trees grown tall," failed to convey the temporal measurement function of "拱." Similarly, in translating "丁忧" (official mourning leave), both models translated it as "mourning for parents" and "filial piety leave," respectively, neither capturing the institutional detail of "officials suspending duties for 27 months of mourning," leading Japanese audiences in an NHK interview to mistakenly believe "丁忧" was a common family mourning ritual (Zhao Tao, 2021). This cognitive gap in folk culture creates "meaning voids" in cross-cultural communication. Especially when the documentary visuals show scenes of "planting trees at graves," the cultural connection between text and image is severed. Viewers see the graves and trees on screen, yet fail to obtain sufficient information from the subtitles to understand the cultural significance of these visual elements, thus diminishing both the documentary's educational value and the viewing experience.

V. OPTIMIZATION PATHWAYS AND FUTURE PROSPECTS

5.1 Technical Level Improvement Suggestions

5.1.1 Construction of Culture-Specific Corpora

It is recommended to build specialized corpora containing:

(1) Cultural Tagging System: Label each vocabulary item with dimensions like "history/literature/religion."

(2) Multi-version Comparison: Include versions with different strategies (literal/free/annotated).

(3) Contextual Knowledge Base: Add background explanations for historical events and literary allusions.

5.1.2 Construction of Dynamic Cultural Knowledge Bases

It is recommended to build dynamic cultural knowledge bases beyond traditional corpora, including:

(1) Cultural Context Maps: For example, link "An-Shi Rebellion" with related historical concepts like "military governor system," "secession of military provinces," and "Tianbao crisis" to form semantic networks.

(2) Cross-Media Annotations: Supplement expert interpretations from programs like Chinese Poetry Conference for metaphors like "朱门酒肉臭," aiding model understanding of metaphorical layers.

(3) Feedback Iteration Mechanism: Collect logs of overseas audience queries for terms like "Taoist immortal" and "imperial examination," automatically updating knowledge base weights. For instance, if the query rate for mistranslations of "黄泉" exceeds 15%, the system automatically triggers a term correction process, supplementing the explanation of "Yellow Spring" as "Chinese underworld, distinct from Western Hades." According to experimental data from the BBC Natural History Unit, dynamic knowledge bases can improve the translation accuracy of cultural terms by 28%, with significant effects especially for multi-dimensional concepts like "24 Solar Terms" and "Five Elements theory."

5.2 Operational Paradigms for Application Scenarios

5.2.1 Human-AI Collaborative Workflow Design

It is recommended to adopt a three-level workflow: "AI Initial Translation - Translator Polishing - Cultural Expert Review."

1. AI completes 70% of standardized translation: Handle non-culture-loaded content like historical events ("An-Shi Rebellion" uniformly translated as "An-Shi Rebellion") and time/place ("天宝三载" automatically generated as "the 3rd year of Tianbao Era"), leveraging DeepSeek's terminology consistency advantage to improve efficiency.

2. Translators handle culture-loaded words: For poetic metaphors like "朱门酒肉臭," combine Moonshot AI's contextual inference results (e.g., "the rich man's gate reeks of wine"), and supplement

annotations explaining the class symbolism of "vermilion gates " (e.g., "vermilion gates symbolize aristocratic households ").

3. Cultural scholars verify conceptual accuracy: Focus on verifying accuracy for easily mistranslated scenarios like "世外道仙 " (needs to distinguish "Taoist immortals " from ordinary hermits) and "墓木已拱 " (supplement cultural interpretation like "arched trees indicate long passage of time "). Taking the translation of "醉眠秋共被 " (Drunk, we slept under one quilt in autumn) from China Season 2 as an example: DeepSeek's initial translation "Drunk, we slept under one quilt in autumn " could be polished by a translator to "Drunk with autumn's charm, we shared a quilt through the night, " preserving the poetic sense of "autumn " while supplementing seasonal imagery through "charm. " The cultural expert's final review must confirm that "共被 " is not misinterpreted as modern intimacy, ensuring it aligns with the cultural context of Tang Dynasty literati interactions.

VI. RESEARCH CONCLUSIONS AND PROSPECTS

6.1 Core Findings and Innovative Value

This study, through systematic comparative evaluation, provides an in-depth revelation of the performance characteristics and limitations of DeepSeek and Moonshot AI in cultural subtitle translation. The main findings can be summarized into the following three levels:

At the technical performance level, the two models exhibit distinct strategic differentiation. DeepSeek excels in preserving cultural symbols, particularly in the literal translation of calendrical terms (e.g., "sui/si/zai "), achieving a cultural symbol retention rate as high as 89%. Its translation results maximally maintain the cultural characteristics of the source language. Moonshot AI, on the other hand, places greater emphasis on the acceptability of the translation. By employing liberal translation strategies, it enhances target language readers' comprehension by 22%, albeit at the cost of some simplification or generalization in conveying culture-specific items. This divergence reflects the different design orientations of the two models: DeepSeek leans towards a "foreignization " strategy, emphasizing cultural fidelity, while Moonshot AI tends towards a "domestication " strategy, focusing on reader acceptance.

At the literary translation level, the study finds that AI exhibits significant "loss of artistic conception " (意境损耗) when handling classical poetry. Taking the translation of "朱门酒肉臭 " as an example, DeepSeek's version ("Behind vermilion gates meat and wine rot ") preserves the visual imagery of the original line (ROUGE-2 score 0.41) but lacks in emotional conveyance. Moonshot AI's version ("The rich man's gate reeks of wine and meat ") strengthens the critical tone through the word "reeks " (BLEU-4 score 0.38) but loses the class symbolism of "vermilion gates " (朱门). This dilemma of being "unable to simultaneously capture form and spirit " (形神难兼) exposes a deep-seated bottleneck in AI's literary translation—a lack of profound understanding of human emotions and aesthetic experiences.

At the cultural cognition level, the study finds that 14.7% of mistranslations are concentrated in cultural concepts requiring background knowledge (e.g., "道仙 " - Taoist immortal, "拱木 " - arched trees, signifying the passage of many years after death). This indicates that current large language models have not yet truly constructed a cognitive schema of Chinese culture. The models can only perform surface-level semantic matching and cannot understand the historical context and philosophical connotations behind cultural symbols. This finding offers an important insight for AI cultural translation research: merely expanding model parameters and training data cannot solve the problem of cultural understanding; it requires the construction of specialized cultural knowledge graphs and cognitive reasoning mechanisms.

The theoretical innovation of this study is mainly reflected in three aspects: First, it constructs a "three-dimensional, nine-category " cultural element analysis framework, achieving refinement and systematization in cultural translation research. Second, it develops a composite "Cultural Adaptation Index (CAI) ", integrating multidimensional data such as BLEU-4 scores (0.29-0.47) and cultural fidelity scores (2.1-3.8 points), providing a quantifiable tool for assessing AI cultural translation. Third, it reveals the dual problems of "technical normative defects " (punctuation error rate 23%)

and "insufficient cultural interpretation " (metaphor misinterpretation rate 31%), pointing out specific directions for subsequent model optimization.

6.2 Limitations and Future Research Directions

This study has the following three main limitations:

In terms of corpus scope, only the first episode of the second season of China was selected as the analysis sample. Although this episode contains rich cultural elements, it does not cover the entire content of the documentary (e.g., themes related to Song Dynasty literati, the spread of Buddhism) nor does it represent the linguistic features and cultural contexts of different historical periods. Future research could expand the corpus to the entire China series (including the pre-Qin and Qin-Han content of the first season) and compare different documentary genres (e.g., the food culture in A Bite of China, the terminology of cultural relics in Every Treasure Tells a Story) to test the universality of the evaluation framework.

In terms of evaluation methods, although this study adopted a hybrid "automation + manual " evaluation model, the team of cultural scholars for annotation consisted of only three people, which may introduce subjective bias. Particularly for more subjective indicators like "cultural fidelity " and "acceptability ", evaluators from different backgrounds might render different judgments. Subsequent research could employ the Delphi method to form a consensus evaluation standard through multiple rounds of expert consultation, while also expanding the size of the annotation team (including sinologists, translation scholars, film and media studies scholars, etc.) to enhance the objectivity and authority of the evaluation results.

In terms of technical comparison, this study focused only on two domestic models, DeepSeek and Moonshot AI, and did not include international mainstream models like GPT-4, Claude, or Gemini for horizontal comparison, making it difficult to comprehensively position the technical coordinates and development level of domestic models. The next step in research could expand the range of models, using controlled variable experiments to compare the advantages of different models in cultural translation (e.g., GPT-4's advantage in long-text reasoning, Claude's strength in linguistic fluency), providing a reference for the targeted optimization of domestic models.

Based on the above limitations, future research could explore the following directions:

First, conduct multimedia translation research. Explore the integration of multiple sources of information from documentaries—such as visuals, sound effects, and subtitles—to build a cross-modal cultural translation evaluation framework. For example, when the screen shows war scenes from the "安史之乱 " (An Lushan Rebellion), the AI could automatically adjust the emotional intensity of the translation, using more impactful vocabulary; when the background music changes to a mournful melody, the translation could correspondingly adjust the sentence rhythm to achieve audio-text synchronized emotional conveyance.Second, strengthen exploration into translation for less-commonly taught languages. This study focused only on Chinese-English translation, but the target audience for promoting Chinese culture internationally encompasses many non-English speaking countries. Subsequent research could apply the evaluation framework to languages such as Spanish, Arabic, and Russian, investigating the commonalities and differences in how AI handles the cultural characteristics of different languages. For instance, the presence of numerous religious expressions in Arabic might lead to significantly different translation strategies for culture-loaded terms compared to English. These comparative findings would greatly enrich cross-cultural translation theory.Third, promote the practical application of human-AI collaboration models. Based on the technical defects and cultural mistranslation problems identified in this study, it is recommended to develop a three-level collaborative process: "AI initial translation - professional translator proofreading - cultural expert review. " The AI would be responsible for basic translation and maintaining terminology consistency, translators would focus on handling culture-loaded terms and literary expressions, and cultural experts would conduct the final review of deep cultural concepts. This model could leverage the efficiency advantages of AI while ensuring the accuracy and depth of cultural translation, truly achieving the communication mission of " helping the world understand China's stories. "

Fourth, construct a dynamic cultural knowledge graph and contextual reasoning mechanism.Current large language models exhibit significant shortcomings in cultural common sense and historical logical reasoning. Future efforts should prioritize the development of a cultural knowledge graph that integrates historical events, social customs, and philosophical concepts, complemented by the introduction of a contextual reasoning module. For instance, when translating "丁忧" (dingyou), the system should not only accurately render it as "mourning leave" but also associate its institutional background (a 27-month mourning period, mandatory leave for officials, etc.) and automatically generate appropriate cultural annotations to help international audiences understand the underlying traditional rites and rituals.

Fifth, develop an AI translation optimization system for real-time scenarios. Addressing the high demands for timeliness and adaptability in documentary subtitle translation, it is necessary to research and develop an adaptive translation framework with online learning capabilities. This system should be able to adjust translation strategies in real-time based on viewer feedback and post-editing results. Furthermore, by continuously learning the stylistic preferences of specific directors or production companies, it can gradually enhance output consistency and cultural adaptability. For example, when processing the China documentary series, the system could learn its poetic linguistic style and automatically apply similar expressions when translating new episodes.

Sixth, establish an interdisciplinary collaborative cultural translation laboratory. It is recommended to form a collaborative team encompassing computational linguistics, classical Chinese literature, film and media studies, and artificial intelligence to jointly build a large-scale experimental platform for Chinese cultural translation. This platform would integrate multimodal cultural resources (such as images of cultural relics, audio recordings of classic texts, and historical maps), allowing researchers to systematically investigate the interaction between language, sound, and imagery in cultural transmission. It would also provide high-quality, multi-dimensional, meticulously culturally annotated Chinese-English parallel corpora for AI model training, fundamentally enhancing models' deep understanding and creative expressive capabilities regarding Chinese culture.

This study provides a solid empirical reference for cultural translation in the age of artificial intelligence, demonstrating both the possibilities brought by technological progress and the long-term nature of the cultural understanding gap. Just as the documentary China reveals—the spirit of the High Tang lies not only in the prosperity of its golden age but also in the cultural perseverance during times of turmoil—the development of AI translation technology also needs to find a balance between the pursuit of efficiency and cultural heritage. Only through the dual drive of technological optimization and humanistic 关怀 (care/consideration) can we truly achieve deep dialogue in civilizational exchange and effectively promote Chinese culture globally.

REFERENCES

[1] Bittner H. The quality of translation in subtitling[J]. Trans-kom, 2011, 4(1): 76-87.

[2] Chuang Y T. Studying subtitle translation from a multi-modal approach[J]. Babel, 2006, 52(4): 372-383.

[3] Cintas J D. Subtitling: Theory, practice and research[M]//The Routledge handbook of translation studies. Routledge, 2013: 273-287.

[4] Ge Jingmin. The Current Status and Development Direction of Domestic Translation Software[J]. Chinese Science & Technology Translators Journal, 2000, (01):31-33. Guo Ningning, Jiang Linhua. Multimodal Video Caption Processing Based on Hard Attention Mechanism[J]. Application Research of Computers, 2021, 38(03):956-960.

[5] Hou Jingyi, Qi Yayun, Wu Xinxiao, et al. Cross-Language Knowledge Distillation for Chinese Video Caption Generation[J]. Chinese Journal of Computers, 2021, 44(09):1907-1921.

[6] Huang, J., & Wang, J. (2023). Post-editing machine translated subtitles: examining the effects of non-verbal input on student translators' effort. Perspectives, 31(4), 620-640.

[7] Hu, K., O'Brien, S., & Kenny, D. (2020). A reception study of machine translated subtitles for MOOCs. Perspectives, 28(4), 521-538.

[8] Koglin, A., Moura, W. H. C., De Matos, M. A., & da Silveira, J. G. P. (2023). Quality assessment of machine-translated post-edited subtitles: an analysis of Brazilian translators' perceptions. Linguistica Antverpiensia,

New Series–Themes in Translation Studies, 22.

[9] Li Yingjun. Current Status and Prospects of Machine Translation and Translation Technology Research: An Interview with Bernard Mark Shuttleworth[J]. Chinese Science & Technology Translators Journal, 2014, 27(01):24-27.

[10] Molina, L., & Hurtado Albir, A. (2002). Translation techniques revisited: A dynamic and functionalist approach. Meta, 47(4), 498-512.

[11] O'Sullivan C. "New and improved subtitle translation ": representing translation in film paratexts[M]//Linguistic and cultural representation in audiovisual translation. Routledge, 2018: 265-279.

[12] Striuk, A. M., & Hordiienko, V. V. (2025). Research and development of a subtitle management system using artificial intelligence. In CEUR Workshop Proceedings (pp. 415-427).

[13] Shen Mengfei, Huang Wei. Tracking Machine Translation Traces: A Study on Syntactic Features of Neural Machine Translation Texts[J]. Foreign Language Teaching and Research, 2024, 56(03):429-441+480-481.

[14] Si Ruo, Yu Yahui. Contemporary Expression in Chinese Historical Documentaries: Taking China as an Example[J]. Sichuan Drama, 2025, (05):32-36

[15] Volk, M., Nivre, J., Dahllöf, M., & Megyesi, B. (2008). The automatic translation of film subtitles: a machine translation success story?. Studia Linguistica Upsaliensia, 202-214.

[16] Volk, M., Sennrich, R., Hardmeier, C., & Tidström, F. (2010). Machine translation of TV subtitles for large scale production. In JEC 2010; November 4th, 2010; Denver, CO, USA (pp. 53-62). Association for Machine Translation in the Americas.

[17] Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in machine translation. Engineering, 18, 143-153.

[18] Wang Huashu, Li Ying. Research on Subtitle Translation Technology: Status, Problems, and Suggestions[J]. Foreign Language World, 2020, (06):80-85+6.

[19] Wang Junsong, Zhuang Congqian, Wei Yongpeng. Machine Translation Quality Evaluation: Methods, Applications, and Prospects[J]. Foreign Language and Literature, 2024, 40(03):135-144.

[20] Yan Xin, Chen Ruizhe, Zhang Jing. Development Status and Trends of Cloud Platforms for Translation Technology[J]. Chinese Science & Technology Translators Journal, 2019, 32(01):22-25.

[21] Yang Yanxia, Wang Yuting, Xiang Yu. Research on Factors Influencing Machine Translation Quality: Evidence from a Meta-Analysis[J]. Foreign Language Research, 2025, (03):26-32. [11].

[22] Zhao Junfeng, Li Xiang. Research on the Construction and Application of Large Language Model-Driven Translation Agents[J]. Foreign Language World, 2024, (05):22-28+75+108.

[23] Zhao Tao. Current Status and Problems of Machine Translation Post-Editing[J]. Foreign Language Education, 2021, 42(04):100-104.

[24] Zainudin, I. S., & Awal, N. M. (2012). Translation techniques: Problems and solutions. Procedia-Social and Behavioral Sciences, 59, 328-334.

[25] Zhou, D., Truran, M., Brailsford, T., Wade, V., & Ashman, H. (2012). Translation techniques in cross-language information retrieval. ACM Computing Surveys (CSUR), 45(1), 1-44.