International Journal on Studies in English Language and Literature (IJSELL)
Volume 13, Issue 9, September 2025, PP 1-11
ISSN 2347-3126 (Print) & ISSN 2347-3134(Online)
https://doi.org/10.20431/2347-3134.1309001
www.arcjournals.org

Assessing the Quality of Artificial Intelligence Translations of Textile Intangible Cultural Heritage Texts: A Big-Data Approach

Yuxuan Zeng[1], Shan Liu[1*]

[1]*School of Foreign Languages, Huazhong University of Science and Technology, Wuhan 430074, China*

Corresponding Author:** ***Shan Liu, *School of Foreign Languages, Wuhan Textile University, Wuhan 430200, China; Institute of Language Sciences, Shanghai International Studies University, Shanghai 200083, China.*

Abstract: *This study compares ChatGPT-4's translation of textile intangible cultural heritage texts with human translation to assess its quality. Using Coh-Metrix 3.0, the study evaluates the translation features, style, and quality, while automatic machine translation metrics like BLEU and TER are employed for quantitative assessment. Qualitative analysis of translation cases highlights key differences between ChatGPT-4 and human translations. The findings show that ChatGPT-4's translation of textile intangible cultural heritage texts is of low quality, characterized by literal translations, lack of cultural sensitivity, and inaccurate use of textile terminology. ChatGPT-4 also struggles with the structure of the original texts. The study emphasizes the importance of accurate translation for the preservation and sustainable dissemination of textile intangible cultural heritage. Suggestions for improving post-editing of ChatGPT-4 translations are provided. This research has practical significance for enhancing China's cultural presence and promoting sustainable development through the dissemination of textile heritage.*

Keywords: *big data; translation quality assessment; textile intangible cultural heritage; artificial intelligence; Coh-Metrix*

1. Introduction

World Heritage refers to the rare and irreplaceable wealth of mankind recognized by UNESCO and the World Heritage Committee. It can be divided into material and intangible cultural heritage, of which Intangible Cultural Heritage refers to oral traditions, folk activities and ceremonial festivals, traditional crafts, and the cultural spaces associated with them. It embodies the thoughts, talents, and emotions of human beings, and has rich connotations. It is the cultural lifeblood of human beings to continue, the precious wealth of world civilization, and has immeasurable multiple values.

Chinese textile intangible cultural heritage, an integral part of world cultural heritage, epitomizes the long history and splendid civilization of the Chinese nation. Inheriting and protecting textile intangible cultural heritage equates to inheriting and protecting history, contributing to sustainable development of intangible cultural heritage. However, due to the influence of foreign cultures, the dissemination of textile intangible cultural heritage is confronting a significant challenge. Language barrier during the process of international dissemination could be resolved through translation. A high-quality translation helps popularizing Chinese textile intangible cultural heritage, telling China's stories well and enhancing China's cultural soft power, which is of paramount importance nowadays.

With the fast development of artificial intelligence, large language models boomed. Unlike traditional machine translation systems, which are trained in sentence units, large language models are trained in word units. This allows the large language model to understand and reproduce the coherence and context information between words, thus making the translation more natural and accurate (Hu & Li, 2023:66). Nowadays, the large language model ChatGPT-4 and generative artificial intelligence have aroused wide attention and heated discussion in the society. Therefore, it is crucial to investigate the performance of ChatGPT-4 in translating texts related to textile intangible cultural heritage.

To gain a more in-depth understanding of the performance of ChatGPT-4 in translating texts related to textile intangible cultural heritage, this study focuses on the following research questions: (1) What's the quality of textile intangible cultural heritage texts translated by ChatGPT-4 as a translation tool? (2)

What are the differences of human translation and ChatGPT-4 translation concerning textile intangible cultural heritage texts? (3) What are the weaknesses of ChatGPT-4 in translating textile intangible cultural heritage texts?

2. Literature Review

2.1. Translation Quality Assessment

Translation quality assessment in translation studies is bound up with theories of translation itself. In the book *Toward a Science of Translating* published in 1964, Eugene A Nida first put forward the concept of "dynamic equivalence", which is also called "functional equivalence." In his perspective, a good translation should be the one that the psychological reactions of readers in the receptor language should be very similar to those of the readers in the source language. House's "translation quality assessment model" (1997) is the world's leading model. Based on the "equivalence" concept and various linguistic application theories, House's assessment model draws conceptual resources from pragmatics, registers, stylistics, discourse analysis theory, especially Halliday's systemic-functional linguistics and systematically integrates them into a framework, attempting to analyze and compare the language-discourse and context-cultural features of the original text and the target text. The model uses parameters to analyze and compare the original text and the translation to look for differences between the two, identifying "mismatches" or "errors" in terms of genre and context.

2.2. Machine Translation Quality Assessment

Machine translation refers to the process of converting linguistic signs from one language to another using computers. Machine translation dates back 1930s and has a history over 90 years. In general, it is divided into four stages, namely rule-based machine translation, statistical machine translation, example-based machine translation and machine translation based on the application of different methods (Hu & Li, 2016:10). With the development of technology, scholars from the field of translation studies began to investigate how to translate languages using computers.

Studies mainly focus on review of machine translation quality assessment and methodologies. Castilho et al. (2018) summarized approaches to assess machine translation quality. Maučec & Donaj (2019) focused on black-box evaluation, one of the two traditional paradigms of machine translation evaluation. Trigueros (2022) also carried out a systemic literature review on machine translation quality assessment.

Despite considerable researches on evaluation of machine translation systems, the field still lacks a commonly accepted standard procedure. Hovy et al. (2002) defined a framework for Machine Translation Evaluation which relates the quality model to the purpose and context of the system. Padó et al. (2009) proposed a metric that assesses the quality of machine translation output through its semantic equivalence to the reference translation.

2.3. AI Translation Quality Assessment

In the 21st century, artificial intelligence has become an important area of research in virtually all fields, including translation studies. ChatGPT, a generative artificial intelligence model created by OpenAI on November 30, 2022, has drawn huge attention from all over the world for its capability of handling language understanding and generation tasks. Recent trends in artificial intelligence have led to a proliferation of studies that focus on the assessment of ChatGPT translation quality.

Studies mainly centered on two aspects, which are comparative analysis of ChatGPT translation quality with that of other translation engines or humans and investigation into characteristics of ChatGPT translation. Scholars tend to compare the quality of ChatGPT translation with that of other translation engines or humans. Khoshafah (2023) evaluates the translation accuracy of ChatGPT3.5 by comparing its outputs with professional translations of various text genres. Sahari et al. (2024) assessed the ability and the effectiveness of ChatGPT in translating figurative language by comparing it with human translations. Several studies focus on the analysis of characteristics of ChatGPT translation. Işım & Balcıoğlu (2023) measured the translation performance of artificial intelligence from English to Turkish. Hendy et al. (2023) also presented a comprehensive evaluation of GPT models for machine translation.

3. METHODOLOGY

3.1. Research Design and Corpus Construction

For the research, the author created a parallel corpus of textile intangible cultural heritage texts from bilingual books titled *Chinese Arts and Crafts* and online resources. The two versions of this book were scanned and kept as PDF forms. ABBYYFineReader was used to transform the PDF form into TXT form. Data cleaning was performed using regular expressions in Microsoft Word. Tokenization and annotation of the Chinese text were conducted via CorpusWordParser, with annotation of the English text implemented through TreeTagger.

In this study, the Chinese version of *Chinese Arts and Crafts* was divided into 20 segments and translated by ChatGPT-4 sequentially. The translation prompt employed was as follows: “Translate the text below in academic English. The text is from a book named ‘Chinese Arts and Crafts’. Readers of this book are not experts in this field. The answer should be shown without quotation marks. The text is: ......” Translations of these 20 segments were integrated into a TXT file in an organized way.

ChatGPT-4 translation is aligned with human translation at sentence level. Alignment was performed via the online platform Tmxmall, accessible at https://www.tmxmall.com/.

Parameters of the three self-built corpora are presented in the following tables:

Table 1. *Parameters of the Chinese corpus (Data from Wordsmith Tools 4.0).*

Parameters	Chinese Corpus
Tokens	54630
Types	1252
Standardized TTR	14.31%

Table 2. *Parameters of human translation corpus (Data from Wordsmith Tools 4.0).*

Parameters	Human translation Corpus
Tokens	40505
Types	5185
Standardized TTR	43.41%

Table 3. *Parameters of ChatGPT-4 translation corpus (Data from Wordsmith Tools 4.0).*

Parameters	ChatGPT-4 translation Corpus
Tokens	34239
Types	5028
Standardized TTR	45.32%

3.2. Instrument

3.2.1. Coh-Metrix 3.0

Coh-Metrix is a computational tool developed by scholars from University of Memphis. According to Graesser et al. (2004), Coh-Metrix is a text analysis system that generates a total of 108 metrics. These indices could be practical to scale a text on characteristics linked to words, sentences and connections between sentences, covering 11 aspects: descriptive, text easability principal component scores, referential cohesion, latent semantic analysis, lexical diversity, connectives, situation model, syntactic complexity, syntactic pattern density, word information, and readability (McNamara et al., 2014).

3.2.2. Automatic Evaluation Metrics

BLEU was proposed by Papineni et al. in 2002. It evaluates the precision of the translation by calculating the n-gram overlap between the machine translation and a set of reference translations. BLEU scores range from 0 to 1. The higher the score is, the better the candidate translation is. TER measures the amount of editing that a human would have to perform to change a candidate reference so it exactly matches a reference translation. The score is between 0 and 1. The higher the score is, the lower the quality is.

In the present investigation, Python scripts were devised to compute the two quality assessment metrics for both the ChatGPT-4 translation and the English rendition of the book *Chinese Arts and Crafts*.

Additionally, three sentences were chosen and their metrics were calculated to facilitate a more nuanced qualitative analysis.

4. Results and Discussion

4.1. Comparative Analysis of Linguistic Features in ChatGPT-4 and Human Translations

Linguistic features of ChatGPT-4 translation are analyzed through results given by Coh-Metrix 3.0. The detailed data are shown below:

Table 4. *Descriptive Indices.*

Text ID	ChatGPT-4 Translation	Human Translation
DESPC	1408	1260
DESSC	1418	1715
DESWC	33793	39786
DESPL	1.007	1.361
DESPLd	0.084	0.68
DESSL	24.069	23.56
DESSLd	12.491	13.477
DESWLsy	1.746	1.642
DESWLsyd	0.971	0.934
DESWLlt	5.406	5.064
DESWLltd	2.792	2.728

Table 5. *Text Easability Principal Component Scores.*

Text ID	ChatGPT-4 Translation	Human Translation
PCNARZ	-1.608	-1.474
PCNARp	5.48	7.08
PCSYNz	0.313	0.19
PCSYNp	62.17	57.53
PCCNCz	1.226	0.906
PCCNCp	88.88	81.59
PCREFz	0	0.06
PCREFp	50	51.99
PCDCz	-0.088	0.435
PCDCp	46.81	66.64
PCVERBz	-1.953	-0.514
PCVERBp	2.56	30.5
PCCONNz	-3.664	-2.895
PCCONNp	0	0.19
PCTEMPz	0.1	-0.068
PCTEMPp	53.98	47.61

Table 6. *Referential Cohesion.*

Text ID	ChatGPT-4 Translation	Human Translation
CRFNO1	0.615	0.53
CRFAO1	0.667	0.585
CRFSO1	0.727	0.637
CRFNOa	0.454	0.361
CRFAOa	0.509	0.422
CRFSOa	0.572	0.475
CRFCWO1	0.112	0.103
CRFCWO1d	0.104	0.109
CRFCWOa	0.074	0.065
CRFCWOad	0.091	0.091
CRFANP1	0.102	0.142
CRFANPa	0.015	0.021

Table 7. *Latent Semantic Analysis.*

Text ID	ChatGPT-4 Translation	Human Translation
LSASS1	0.332	0.285

LSASS1d	0.183	0.192
LSASSp	0.229	0.26
LSASSpd	0.149	0.178
LSAPP1	0.334	0.327
LSAPP1d	0.183	0.191
LSAGN	0.401	0.38
LSAGNd	0.098	0.103

Table 8. *Lexical Diversity.*

Text ID	ChatGPT-4 Translation	Human Translation
LDTTRc	0.257	0.252
LDTTRa	0.15	0.131
LDMTLD	80.12	69.267
LDVOCD	119.331	100.012

Table 9. *Connectives.*

Text ID	ChatGPT-4 Translation	Human Translation
CNCAll	103.808	100.865
CNCCaus	19.649	25.537
CNCLogic	23.733	26.693
CNCADC	8.108	9.425
CNCTemp	15.625	18.901
CNCTempx	14.53	14.829
CNCAdd	70.695	59.719
CNCPos	96.677	94.003
CNCNeg	5.948	6.636

Table 10. *Situation Model.*

Text ID	ChatGPT-4 Translation	Human Translation
SMCAUSv	25.36	24.858
SMCAUSvp	32.551	35.993
SMINTEp	9.558	9.576
SMCAUSr	0.283	0.447
SMINTEr	1.738	2.343
SMCAUSlsa	0.043	0.07
SMCAUSwn	0.406	0.513
SMTEMP	0.865	0.832

Table 11. *Syntactic Complexity.*

Text ID	ChatGPT-4 Translation	Human Translation
SYNLE	6.805	7.247
SYNNP	1.304	1.213
SYNMEDpos	0.668	0.629
SYNMEDwrd	0.907	0.879
SYNMEDlem	0.896	0.858
SYNSTRUTa	0.102	0.097
SYNSTRUTt	0.086	0.084

Table 12. *Syntactic Pattern Density.*

Text ID	ChatGPT-4 Translation	Human Translation
DRNP	372.651	373.222
DRVP	136.922	151.536
DRAP	23.91	23.878
DRPP	128.784	139.798
DRPVAL	8.493	14.754
DRNEG	1.923	2.438
DRGERUND	38.026	26.567
DRINF	6.895	10.431

Table 13. *Word Information.*

Text ID	ChatGPT-4 Translation	Human Translation
WRDNOUN	369.1	339.668
WRDVERB	119.02	115.97
WRDADJ	121.268	106.746
WRDADV	36.604	37.601
WRDPRO	12.725	15.86
WRDPRP1s	0.089	0.05
WRDPRP1p	0.089	0.251
WRDPRP2	0.059	0.151
WRDPRP3s	3.344	4.901
WRDPRP3p	4.054	4.373
WRDFRQc	1.777	1.901
WRDFRQa	2.669	2.854
WRDFRQmc	0.655	0.747
WRDAOAc	371.572	361.215
WRDFAMc	543.079	548.305
WRDCNCc	424.768	413.385
WRDIMGc	441.638	433.253
WRDMEAc	436.379	434.909
WRDPOLc	3.36	3.819
WRDHYPn	6.497	6.513
WRDHYPv	1.799	1.674
WRDHYPnv	2.364	2.251

Table 14. *Readability*

Text ID	ChatGPT-4 Translation	Human Translation
RDFRE	34.935	44.375
RDFKGL	14.307	12.833
RDL2	6.539	8.503

4.1.1. Lexical Measures

In Table 4, as for ChatGPT-4 translation, the score of word number is lower and scores of mean number of syllables and letters in words are higher. Table 8 indicates that the type-token ratio of ChatGPT-4 translations is higher than that of human translations, suggesting that ChatGPT-4 demonstrates higher lexical diversity than human translators when translating textile intangible cultural heritage texts.

From the first five indices of Table 13, it can be observed that ChatGPT-4 employs more nouns, verbs, and adjectives, and less adverbs and personal pronouns, compared to human translators. Given that the scores for average word frequency of content words, average word frequency of all words, and average minimum word frequency in sentences of ChatGPT-4 translations are lower than those of human translations, it can be inferred that ChatGPT-4 employs less content words, demonstrating greater lexical diversity. While the score of age of acquisition of ChatGPT-4 translation is higher, that of familiarity is lower than human translation, indicating that ChatGPT-4 uses easy but unfamiliar words. The concreteness, imageability and meaningfulness scores of ChatGPT-4 translations are higher and scores of polysemy and hypernymy for nouns are lower, corresponding to the previous result of word concreteness in Text Easability Principal Component Scores.

4.1.2. Syntactic Measures

Table 11 indicates that ChatGPT-4 translation has a more complex syntactic structure than human translation in general. For instance, mean number of modifiers per noun-phrase, mean minimum editorial distance score between adjacent sentences computed from part of speech tags, words and lemmas, proportion of intersection tree nodes of ChatGPT-4 translation are higher than those of human translation. An exceptional index is SYNLE, which refers to the mean number of words before the main verb of the main clause in sentences. In Table 12, incidence scores of adverbial phrases and gerunds of ChatGPT-4 translation are higher than those of human translation. It could be inferred that ChatGPT-4 uses more adverbial phrases and gerunds.

4.1.3. Cohesion Measures

Table 6 indicates that the scores for co-reference noun overlap, argument overlap, and content word overlap in ChatGPT-4's translations are higher than those in human translations. An exceptional index is anaphor overlap, where both local anaphor overlap and global anaphor overlap of ChatGPT-4 are lower. It could be observed from Table 5 that principal component scores of deep cohesion and verb cohesions of ChatGPT-4 translation are lower than those of human translation. Table 9 shows that incidence of all connectives of ChatGPT-4 translation excels that of human translation. In ChatGPT-4 translation, there are more additive connectives and positive connectives while the incidence of other connectives including causal, logic, adversative, temporal, extended temporal and negative are lower than in human translation. It should be concluded from the above analysis that although ChatGPT-4 translation demonstrates a higher score of referential cohesion in general, its principal component scores of deep cohesion and verb cohesions are lower.

4.1.4. Difficulty Measures

The easability components provided by Coh-Metrix provide a complete picture of text ease and difficulty. In Table 5, although ChatGPT-4 translation has simpler syntax, more concrete words and more consistent temporality, its cohesion and narrativity scores are lower than those of human translation.

4.2. Analysis of BLEU and TER Metrics

Results show that the BLEU score is 0.153 and the TER score is 0.927. According to Zhou and Liu (2022), if the BLEU score of a machine translation reaches 31.4%, it indicates that the translation is of good quality, meeting the basic requirements of machine translation. In this study, the BLEU score is just 0.153. Coupled with the high TER score, the result reveals that ChatGPT translation is of low quality.

4.3. Summary of Quantitative Analysis

4.3.1. Vocabulary

Though less words in ChatGPT-4 translation, ChatGPT-4 boasts a more diversified vocabulary than human translators when translating textile intangible cultural heritage texts and uses more concrete words. For some professional textile terms, ChatGPT-4 may fail to recognize them, resulting in various versions of one proper noun in different circumstances.

4.3.2. Syntax

In Table 5, the syntactic simplicity score of ChatGPT-4 translation is higher, while in Table 11, the mean number of modifiers per noun-phrase, mean minimum editorial distance score between adjacent sentences computed from part of speech tags, minimum editorial distance score between adjacent sentences computed from words and lemmas and proportion of intersection tree nodes of ChatGPT-4 translation are higher, revealing that the overall syntactic complexity score of ChatGPT-4 translation is higher than human translation. Therefore, the result is contradictory. This may be because that for the two syntactic simplicity indices in Table 5, word number weighs heavier, while in Table 11, syntax is more important.

4.3.3. Cohesion

The cohesion of ChatGPT-4 translation is inferior to human translation in general. In Chinese version of the book, there are more explicit and repeated phrases. However, English focus more on logic and internal connection. A possible reason is that ChatGPT-4 translates texts word for word, ignoring the underlying connection between words.

4.3.4. Difficulty

ChatGPT-4 translation is more difficult to read than human translation. The longer a sentence is and the more word syllables are, the more difficult the text is. As shown in Table 4, scores of average sentence length and average number of syllables of ChatGPT-4 translation are higher than those of human translation. Therefore, the readability of ChatGPT-4 translations is inferior to that of human translations

4.3.5. Accuracy

An estimation of the accuracy of analysis can be made from the BLEU and TER scores. The BLEU score is 0.153, indicating a low precision of ChatGPT-4 translation, whereas the TER score is 0.927, indicating a high number of discrepancies compared to human translation. The two metrics verify the low translation quality of ChatGPT-4 when translating textile intangible cultural heritage texts.

4.4. Case Studies

For a more detailed analysis of ChatGPT-4 translation quality, the author chose three typical sentences from the corpus. All example sentences are shown below:

Table 15. *Example Sentences.*

E.g. (1) ：染织工艺方面，三国两晋时期，四川生产的蜀锦最为著名。
E.g. (2) ：水力大纺车是中国古代将自然力运用于纺织机械的一项重要发明。
E.g. (3) ：唐代刺绣技法仍沿袭汉代锁绣，但针法已开始以平绣为主，并采用多种针法和色线，所用绣底质料亦不限于锦帛和平绢。

Then the author calculates both BLEU and TER automatic assessment metrics of ChatGPT-4 translation and the results are shown below:

Table 16. *BLEU Score for ChatGPT-4 Translation (Unit: %).*

E.g. (1)	E.g. (2)	E.g. (3)
25.9	15.7	1.1

Table 17. *TER Score for ChatGPT-4 Translation (Unit: %).*

E.g. (1)	E.g. (2)	E.g. (3)
52.9	85.7	94.7

A qualitative analysis will be conducted to the three sentences:

E.g. (1) 染织工艺方面，三国两晋时期，四川生产的蜀锦最为著名。

Human translation: In the respect of dyeing and weaving technology, the figured satin woven in Sichuan was the most famous during the period of Three Kingdoms and the Western Jin Dynasty and the Eastern Jin Dynasty.

ChatGPT-4 translation: In terms of dyeing and weaving, Shu brocade produced in

Sichuan was the most famous during the Three Kingdoms and Jin Dynasties.

In this instance, two distinct discrepancies can be identified between ChatGPT-4's translation and the human translation. "蜀锦" is translated by humans into "figured satin" but "Shu brocade" by ChatGPT-4. Evidently, ChatGPT-4 adopts a literal translation approach without grasping the underlying meaning of the phrase. A second prominent difference lies in the translation of "三国两晋时期". In human translation, it is translated into "the period of Three Kingdoms and the Western Jin Dynasty and the Eastern Jin Dynasty", while in ChatGPT-4 translation, it is "the Three Kingdoms and Jin Dynasties." Evidently, the ChatGPT-4 translation lacks basic knowledge of traditional Chinese dynasties.

E.g. (2) 水力大纺车是中国古代将自然力运用于纺织机械的一项重要发明。

Human translation: It is a significant invention in ancient China using natural forces for textile machinery.

ChatGPT-4 translation: The water-powered large spinning wheel was an important invention in ancient China that applied natural force to textile machinery.

In this example, human translation omits some information and uses "it" to replace it because it has been mentioned in previous texts. Furthermore, "重要发明" is rendered as "an important invention" by ChatGPT-4 and "a significant invention" by human translators. From these two translations, it can be inferred that ChatGPT-4 adopts a literal translation approach and exhibits a tendency to employ more commonly used vocabulary.

E.g. (3) 唐代刺绣技法仍沿袭汉代锁绣，但针法已开始以平绣为主，并采用多种针法和色线，所用绣底质料亦不限于锦帛和平绢。

Human translation: Although the Han-Dynasty locking embroidery technique was still followed in the Tang Dynasty, another skill called plain stitching method had already been widely used together with some other stitching methods, using color thread and wider range of fabrics.

ChatGPT-4 translation: Tang dynasty embroidery continued to use the Han dynasty's lock stitch but began to mainly use flat stitch, employing various stitches and colored threads on a variety of embroidered base materials.

ChatGPT-4 translates the sentence straightforwardly, while humans add "although" and "still" to make the translation more logically acceptable by native speakers. As for the translation of the textile terminology "平绣", ChatGPT-4 just translates it literally and mistranslates it into "flat stitch". The hugely different sentence structure leads to a very low BLEU score and high TER score.

4.5. Summary of Qualitative Analysis

Based on the qualitative analysis in previous text, several conclusions could be made.

First, ChatGPT-4 isn't sensitive to texts strongly related to Chinese textile culture. When encountering professional textile terms, ChatGPT-4 translates part of them or translates texts word for word due to a lack of knowledge of the underlying cultural connotation.

Second, ChatGPT-4 translation fails to achieve the goal of successfully conveying the meaning of textile intangible cultural heritage. The precision of ChatGPT-4 translation is poor due to many mistranslations and transliterations.

Third, when dealing with sentence structure, ChatGPT-4 is too rigid to change the syntax to make it more logically acceptable to English readers.

Though ChatGPT-4 automatically generates content related to Chinese traditional culture according to the input of previous corpus, its translation is based on many English trainings, and most of the target groups in cross-cultural communication are native English speakers, hence the lack of Chinese textile intangible cultural heritage training. Language models such as ChatGPT may produce misleading information, causing the public to form wrong knowledge and value judgments, and even deepening cultural stereotypes and distorting the values of the Chinese nation.

5. Conclusion and Implications

The research finds that ChatGPT-4 has inconsistency problems when translating professional textile terms. Given the particularity of such terms, translations must maintain consistency and precision. However, ChatGPT-4 frequently alters expressions in practical application, which may lead to misunderstandings among English readers. To address this issue, ChatGPT-4 should adopt a more rigorous and unified translation strategy to ensure the clarity and consistency of translated texts.

The study reveals that ChatGPT-4 has significant inadequacies in the knowledge of China's textile intangible cultural heritage. Due to the lack of training data, artificial intelligence faces difficulties in identifying culture-specific terms. Such a lack of cultural cognition not only affects the accuracy of translations but also weakens the its ability to effectively convey the essence of China's textile cultural heritage. Furthermore, ChatGPT-4 fails to provide necessary cultural context. This limitation underscores the crucial necessity of human intervention in fields that require in-depth cultural understanding, such as translation of cultural heritage texts. We also find that ChatGPT-4 tends to generate lengthy sentences, which significantly reduces the readability of the translated text. More importantly, there is a lack of deep semantic connections between words, and the cohesion of translated texts is notably inadequate.

The conclusion indicates that although ChatGPT-4 exhibits promising potential, it still requires extensive human intervention to meet the expected standards in professional translation fields. Post-editors play a key role in improving the quality and accuracy of large language model translation, ensuring that the translated text is both accurate and culturally sensitive. This study not only identifies the limitations of current large language model translation technology but also provides practical suggestions for enhancing the efficiency of technology application in professional translation processes.

FUNDING

This paper is a special research grant program of Wuhan Textile University. Therefore, the funder is Wuhan Textile University and the grant number is 2024476.

REFERENCES

杭间,郭秋惠.中国传统工艺[M].北京:五洲传媒出版社, 2023.

Hang J, Guo Q H. *Chinese Arts and Crafts* [M]. Beijing: China Intercontinental Press, 2023.

胡开宝,李晓倩.大语言模型背景下翻译研究的发展：问题与前景[J].中国翻译,2023(6):64-73.

Hu K B, Li X Q. Development of Translation Studies in the Context of Large Language Models: Issues and Prospects [J]. *Chinese Translators Journal*, 2023(6):64-73.

胡开宝,李翼.机器翻译特征及其与人工翻译关系的研究[J].中国翻译, 2016, 37(5):10-14.

Hu K B, Li Y. Research on the Characteristics of Machine Translation and Its Relationship with Human Translation [J]. *Chinese Translations Journal*, 2016, 37(5):10-14.

周成彬,刘忠宝.基于语义信息共享Transformer的古文机器翻译方法[J].情报工程, 2022,8(6):114-127.

Zhou C B, Liu Z B. Machine Translation of Ancient Chinese Text Based on Transformer of Semantic Information Sharing[J]. *Information Technology*, 2022, 8(6):114-127.

Castilho S., Doherty S., Gaspari, F., et al. Approaches to human and machine translation quality assessment. In *Translation quality assessment: From Principles to Practice*; Moorkens, J., Castilho, S., Eds.; Springer Publishing Company: New York, America, 2018; Volume 1, pp. 9-38.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., et al. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers* 2004, 36, 193-202.

Hendy, A., Abdelrehim, M., Sharaf, A., et al. How good are GPT Models at Machine Translation? A Comprehensive Evaluation. *arXiv* 2023.

House, J. *Translation Quality Assessment: A Model Revisited*; Tübingen: Narr, 1997.

Hovy, E., King, M., Popescu-Belis, A. Principles of Context-Based Machine Translation Evaluation. *Machine Translation* 2002, 17, 43-75.

Işım, Ç. & Balcıoğlu, Y. S. ChatGPT: performance of translate. In *3rd International ACHARAKA Congress on Humanities and Social Sciences Proceedings Book*, 2023; 47-51.

Khoshafah, F. Chatgpt for Arabic-English Translation: Evaluating the accuracy. *Research Square* 2023, 1-19.

Maučec, M. S. & Donaj, G. Machine Translation and the Evaluation of its Quality. In *Recent Trends in Computational Intelligence*; Sadollah, A., Sinha, T. S., Eds; Academic Press: New York, America, 2021; Volume 143, pp.1-20.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., et al. *Automated Evaluation of Text and Discourse with Coh-Metrix*; Cambridge University Press: Cambridge, America, 2014.

Padó, S., Cer, D., Galley, M., et al. Measuring Machine Translation Quality as Semantic Equivalence: A Metric Based on Entailment Features. *Machine Translation* 2009, 23, 181-193.

Papineni, K., Roukos, S., Ward, T., et al. Bleu: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 2002; 311-318.

Rivera-Trigueros, I. Machine Translation Systems and Quality Assessment: A Systematic Review. *Language Resources and Evaluation* 2022, 56, 593-619.

Sahari, Y., Qasem, F., Asiri, E., et al. Evaluating the Translation of Figurative Language: A Comparative Study of ChatGPT and Human Translators. *Research Square* 2024.

AUTHORS' BIOGRAPHY

Yuxuan Zeng, graduate student at School of Foreign Languages, Huazhong University of Science and Technology. Academic interests: second language acquisition.

Shan Liu, associate professor at School of Foreign Languages, supervisor of MTI(Master of Translation and Interpreting) postgraduates, Wuhan Textile University, PhD candidate at Institute of Language Sciences, Shanghai International Studies University. Academic interests: corpus-based translation studies and second language acquisition.

Citation: *Yuxuan Zeng & Shan Liu. "Assessing the Quality of Artificial Intelligence Translations of Textile Intangible Cultural Heritage Texts: A Big-Data Approach". International Journal on Studies in English Language and Literature (IJSELL), vol 13, no. 9, 2025, pp. 1-11. DOI: https://doi.org/10.20431/2347-3134.1309001.*