ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on the Influence Mechanism of Artificial Intelligence Tools on the Translation Quality of Literary Works from the Perspective of Human-Computer Collaboration

Wanling He[1] *and Chunlan Liang*[1,*]

[1] School of Foreign Studies, Guangxi Vocational Normal University, Nanning, 530007, Guangxi, China

SUMMARY: *How to maximize the technical potential of artificial intelligence tools under the premise of fully guaranteeing the translator's subjective status has become the most central and urgent practical issue in the field of contemporary translation studies. This study establishes a systematic theoretical analysis framework, discusses in depth the operation mechanism of human-computer collaborative mode in the translation practice of literary works, and establishes a human-computer collaborative translation quality optimization system with this support. Twenty-four senior undergraduate students were used as research subjects for the translation practice experiment. The results show that the experimental group is significantly better than the control group in all evaluation dimensions, and the comprehensive quality score is improved by 36.8% compared with the control group. The research in this paper not only provides a theoretical basis and empirical data support for translation practice activities in the era of intelligence, but also has very important practical significance for promoting the deep-level reform in the field of translation education and improving the quality of translation personnel training.*

KEYWORDS: *artificial intelligence; human-computer collaboration; literary translation; translation quality; learning investment*

1 Introduction

1.1 Background of the study

In recent years, with the rapid development of natural language processing technology, the artificial intelligence technology represented by ChatGPT has demonstrated impressive language processing capabilities, which has brought new opportunities and challenges to the traditional language service industry, thus arousing extensive attention and discussion in industry and academia [1-4]. In the foreign language learning industry, some experts suggest that although the ChatGPT big language model can create meaningful contexts through cooperative dialogues, thus efficiently empowering second language acquisition, it should also be noted that it may have problems such as "information cocoon" [5]. In the translation industry, compared with the traditional human translation, the application of ChatGPT and other technologies is a low-cost, high-efficiency option that can create a large number of translation results in a short period of time, which is not only used to translate a lot of political, economic, legal, scientific and technological fields of the easier to standardize the

*Liangchunlan2005@126.com
https://doi.org/10.65102/is2026397
Published: 30 April, 2026

"hard text" [6], but also used in the translation of poetry, prose, and other languages. It has also been applied to "soft texts" such as poems, essays, novels, miscellaneous essays, reviews and other literature [7]. Due to the difficulty of finding good translators, there is even a website LNMTL (Light Novel Machine Translations) that specializes in machine translation of Chinese online literature.

In recent years, deep learning has made significant progress in the fields of natural language processing (NLP) and neural machine translation (NMT). From Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) to Long Short-Term Memory Networks (LSTM) to the Transformer Architecture based on the Attention Mechanism, the continuous iteration of the technology has given machine translation brand new opportunities and challenges [8, 9].The emergence of Large Models (PTMs), such as BERT, GPT and so on, has further accelerated the development of this field. The introduction of Large Language Models (LLMs), such as DeepSeek, ChatGPT and Gemini, and the emergence of inference models based on Chain of Thought (CoT) training tools (e.g., DeepSeek R1, OpenAI o3) have revolutionized machine translation [10]. In addition, according to Son and Kim (2023), the global machine translation market size is expected to grow from $810 million in 2022 to $4.07 billion in 2030, at a CAGR of 19.9% [11]. Meanwhile, the construction of new liberal arts has promoted the deep integration of language translation and artificial intelligence [12]. However, the current humanities and social sciences community has paid insufficient attention to the application of machine translation, and most of the research flows from theoretical discussion, with little quantitative analysis and use case dissection, especially in literary translation [13]. Therefore, it is of great practical significance and application value to explore the application prospects and practical effects of emerging AI technologies in literary translation.

1.2 Status of research

In recent years, the research field of machine translation has gradually made important breakthroughs, and the scope of research has been expanding, touching on a variety of fields such as general-purpose translation, low-resource language translation, multilingual translation and so on [14]. The performance of machine translation has also been significantly improved by the rise of large language modeling, which has triggered strong research interest in the academic community. First, Kalchbrenner and Blunsom (2013) proposed an end-to-end (Seq2seq) generative model architecture for machine translation modeling, and began to use RNNs to process machine translation, which also marked the successful birth of neural network-based machine translation [15]. Kalyani and Sajja (2015) conducted a systematic review of machine translation for the rule-based, hybrid, and statistical approaches to translation, presenting measures of the quality of human and machine translation, and the challenges of assessing translation quality [16]. With the continuous progress of machine learning technology, Vaswani et al. (2016) discarded the traditional CNN and RNN architectures and proposed the Transformer model based on a pure attention mechanism, which can deal with long-distance dependencies more efficiently, improve the accuracy and fluency of machine translation to an unprecedented level, and become the new generation of neural network machine translation models the cornerstone of the new generation of neural network machine translation models [17]. Aiming at the problem of Transformer model with many parameters and complex training, Wang et al. (2019) improved the traditional Transformer model by comprehensively using the layer normalization method and Dynamic Layer Linear Combination (DLCL), which can provide thing excellent translation performance with less number of parameters and day training speed is also faster [18]. Thu

(2020) proposed a Transformer model with back-translation, which is more effective for low-resource language translation tasks like English-Vietnamese, and its BLEU scores are improved [19]. Liu and Chen (2022) improved the traditional Transformer model by proposing an X-Transformer based model, which achieved achievable BLEU scores of 46.63 and 55.63 in English-German and English-French translation tasks, which are significantly better than the traditional Transformer [20].

In addition to general-purpose fields, research on translation in specialized fields is progressing rapidly. Scholars have mostly evaluated the effectiveness of AI tools in translating different text types and pointed out the advantages and limitations of AI tools compared with traditional machine translation, mainly related to legal translation, medical text translation, translation of ancient books, translation of literary works, etc. [21-23]. Skianis et al. (2020) evaluated translation quality for statistical machine translation (SMT) and NMT methods used for medical terminology and found that the traditional SMT model performed well when dealing with short sentences of medical text, whereas the NMT model performed better when dealing with long sentences and also demonstrated better performance for ICF (International Classification of Functioning, Disability and Health) terminology [24]. Verma et al. (2022) proposed a novel algorithmic framework for translating ancient texts (Old Sudanese to English), which showed higher translation quality in comparison with professional human translators, with an 8% increase in BLEU scores and a 10% reduction in WER (Word Error Rate) [25]. Zhang et al. (2022) used a deep learning model to compute and count the information in this text of ancient books and realized the automatic translation of the text, and the study found that the sequence model (seq) has the best translation effect, which is able to translate the text of ancient books in a fast-reading and simple way [26]. Alkatheery (2023) found a large number of errors in machine translation in the process of translating Arabic legal texts into English, mainly in terms of vocabulary (43.4%) and errors related to the legal context (30.2%), suggesting that there are still imperfections in the current machine translation in specialized areas [27].

In the field of translation of literary works, Vázquez and Mitkov (2023) used three machine translation systems to evaluate the translation performance of literary works in different genres in order to objectively assess the performance of machine translation systems in each genre [28]. Hu and Li (2023) explored the performance of DeepL in translating literary works (Shakespeare) and compared it with Liang Shiqiu's Chinese translations, and found that DeepL performs well, with more than 80% accuracy and fluency, and shows innovation in translation methods for literary works [29]. Focusing on the challenging task of literary translation, Wu et al. (2024) proposed a multi-intelligence-based virtual company (Trans Agents) for machine translation of literary works and tried to explore two translation quality assessment strategies [30]. Karabayeva and Kalizhanova (2024) pointed out that the current state-of-the-art large language models (ChatGPT and DeepL) show significant advantages in translating literary works, but they still have some limitations and need to be reviewed and improved manually [31]. Moreno and Mora (2025) argued that neural machine translation can provide very reliable performance in translating literary works, and in the task of translating the literary work Don Quixote, DeepL performed better in terms of accuracy and recall, while Google Translate lacked achieved higher scores in terms of BLEU metrics [32].

With the wide application and accelerated iteration of large language models, the methods and indexes for translation quality assessment are becoming more and more diversified. Given that it is difficult for a single index (e.g., BLEU) to fully reflect the translation quality, scholars are committed to exploring other machine translation indexes, which are broadly categorized into three types, namely, lexical level (e.g., BLEU, TER, etc.), unsupervised

learning (e.g., MEANT, BERT Score, BART Score, etc.), and supervised learning (e.g., BEER, NUBIA, COMET, etc.) [33-35]. Lo and Wu (2013) proposed two metrics, MEANT and UMEANT, for assessing the quality of machine translation, which are evaluated by measuring the similarity of semantic frames, with MEANT using a supervised optimization approach and UMEANT using an unsupervised estimation approach, both of which can achieve comparable correlation with human judgments through monolingual corpora and automatic shallow semantic parsers [36]. Zhou and Bollegala (2019) proposed an unsupervised method without labeling data-Bidirectional Minimum Word Movement Distance (BiMWMD)-for evaluating translation quality metrics, and it was found that BiMWMD is a reliable method for evaluating the quality of translations, and its results are highly correlated with the grades rated by professional translators [37]. Mukherjee and Shrivastava (2022) introduced a no-supervision required metric (REUSE) for evaluating translation quality at the sentence and chunk level, which utilizes a multilingual chunker, BERT contextual word embeddings, and a language-independent BERT model that is highly correlated with human judgments [38].

In summary, the current research on machine translation is developing rapidly, the rise of big language models has significantly improved the translation quality and efficiency, the application research of big language models in translation in professional fields is getting more and more attention, the translation quality assessment methods are becoming more and more diversified, and the potential of big language models in translation field in China is highlighted. However, the field still faces many challenges. Firstly, there is a relative lack of research on machine translation of complex texts such as literary works; secondly, most of the research on machine translation focuses on theoretical discussions and lacks quantitative analysis and practical applications; and thirdly, the research on performance optimization of literary works has yet to be in-depth.

2 Methodology

2.1 Experimental design

Based on the concept of quasi-experimental design, this paper constructs a systematic framework for a controlled experiment on translation practice with the aim of exploring the impact mechanism produced by applying the human-computer collaborative translation model in the quality of translation of literary works. Meanwhile, following the principle of scientific research, the experimental subjects were screened through a purposive sampling strategy to select 24 senior undergraduates majoring in translation from the School of Foreign Languages of the university the author is working for. The inclusion criteria are as follows:

(1) Have systematically studied translation theory and practice courses and have solid bilingual skills.

(2) Have no less than one year’s practical experience in literary translation and are proficient in various translation techniques.

(3) Have basic knowledge of intelligent translation tools but have relatively limited experience in using them.

(4) Volunteer to participate in the research activities and be able to complete all the experimental sessions as required.

In order to ensure that the samples are well represented, the selected subjects are concentrated in the age group of 21-23 years old, with a male-to-female ratio of 8:16, and they have the highest English proficiency among senior undergraduates majoring in

translation. All the participants have signed a detailed informed consent form and fully understood the objectives of the experiment before the experiment formally started.

A randomized table method was used to divide the 24 subjects into equal groups, i.e., the experimental group and the control group, with 12 members in each group. The experimental group adopted the human-computer collaborative translation model, which allowed them to obtain systematic intelligent feedback and professional suggestions during the translation process by utilizing mainstream translation platforms such as ChatGPT-4 and DeepL. The control group, on the other hand, followed the traditional manual translation scheme, which only allowed the use of paper dictionaries and related reference materials as auxiliary tools.

A literary fragment of about 500 words from Virginia Woolf's *To the Lighthouse* was chosen for the experimental task, which contains rich modernist literary expression techniques such as metaphor, symbolism and stream of consciousness. The selection of the text strictly follows the three basic principles of moderate linguistic complexity, diversified rhetorical techniques, and deep cultural connotation, which can fully reflect the challenging nature of literary translation without affecting the operability of the experiment due to excessive difficulty. The time limit for the completion of the translation task is set at 120 minutes, which has been fully verified by the pre-experiment to ensure that the participants have sufficient time for in-depth thinking and repeated revisions, and to effectively avoid the fatigue effect caused by too much time. The members of the experimental group can call the AI tool at any time to obtain vocabulary suggestions, grammar checking, stylistic analysis and other aspects of support, and the system automatically records the frequency of interaction with the intelligent tool and the specific way.

The experimental process is executed in strict accordance with standardized procedures, and its specific implementation process is shown in Figure 1.

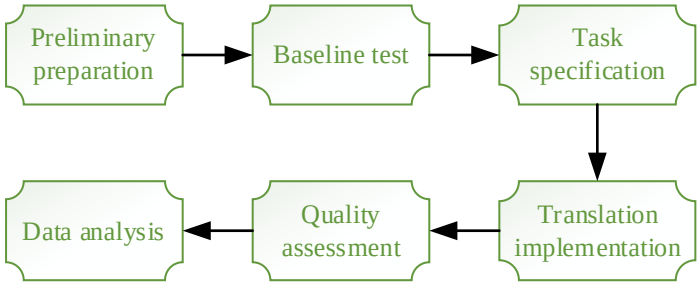

Figure 1: Experimental procedure

The experimental process consists of a total of six key implementation stages, as follows:

(1) Preparation stage. In this phase, the main purpose is to allow participants to fully familiarize themselves with the experimental environment and master the relevant tool operation process, in order to lay the foundation for carrying out the experiment.

(2) Baseline testing phase. This stage is mainly to assess the participants' standardized translation ability, and then fully understand the participants' initial level, so as to establish a reliable baseline for subsequent data analysis.

(3) Task description stage. This stage is mainly to explain the requirements, evaluation criteria and precautions of the translation task, so as to ensure that the translation experimental task can be carried out smoothly.

(4) Translation implementation stage. This stage is the key part of the translation experiment process, which requires participants to complete the translation task in strict accordance with the established grouping, and at the same time, carry out the whole process of supervision and detailed recording of key behavioral data.

(5) Quality assessment stage. This stage is an important link to ensure the specialization of data, relying on the professional evaluation team to implement multi-dimensional comprehensive evaluation of translation results.

(6) Data analysis stage. This stage is the final stage of the comprehensive application of data, mainly through a variety of statistical analysis methods, in-depth analysis of experimental data and systematic testing of research hypotheses.

In order to comprehensively and objectively assess the actual impact of human-computer collaboration on translation quality, this paper constructs an evaluation index system covering multiple dimensions. The following mathematical model is used for quantitative analysis of translation quality assessment:

$$Q_{total} = \alpha \cdot Q_{accurary} + \beta \cdot Q_{fluency} + \gamma \cdot Q_{adequacy} + \delta \cdot Q_{style} \tag{1}$$

where Q_{total} represents the overall quality score of the translated work, $Q_{accurary}, Q_{fluency}, Q_{adequacy}, Q_{style}$ corresponds to the scores of accuracy, fluency, adequacy, and stylistic preservation, respectively, and the respective weighting factor $\alpha, \beta, \gamma, \delta$ are set to 0.3, 0.25, 0.25, 0.2 according to the uniqueness of the literary translation to ensure the scientificity and rationality of the evaluation results.

2.2 Data collection

This study establishes a multivariate data collection framework and examines the completeness and reliability of the research data through standardized procedures and quality control mechanisms. The collection of translation data is mainly through a standardized electronic document submission system, in which the two groups of participants are required to upload their translations within a specified period of time on a dedicated platform, and the system automatically records data such as the submission time, frequency of revisions, word count statistics, and so on. To ensure an objective and fair evaluation process, all translation works are anonymized and independently scored by a number of judges with rich practical experience in literary translation. The evaluation is mainly based on linguistic accuracy, literary retention, cultural communication effect, and overall fluency. The scoring standard is a 10-point scale, and the overall translation quality score Q is calculated through a weighted average formula, i.e.:

$$Q = \frac{1}{m} \sum_{j=1}^{m} [\alpha \cdot A_j + \beta \cdot L_j + \gamma \cdot C_j + \delta \cdot F_j] \tag{2}$$

where m is the total number of reviewers, A_j, L_j, C_j, F_j denote the ratings of the j th reviewer in the dimensions of Accuracy, Literature, Culture, and Fluency, respectively, with the weighting coefficients set to be consistent with the previous section.

The questionnaire survey used a standardized scale designed based on the framework of learning input theory to measure the participants' emotional input, cognitive load, behavioral performance and other dimensions during the translation process, the details of which are shown in Table 1. The questionnaire consisted of three core modules, the Affective Engagement Scale (12 questions), the Cognitive Engagement Scale (10 questions), and the Behavioral Engagement Scale (8 questions), all of which were scored on a 7-point Likert scale ranging from "completely disagree" (1 point) to "completely agree" (7 points). Additional measurement items such as experience of using AI tools, satisfaction with collaborative work, and preference for translation strategies were also added.

Table 1: Structural design of questionnaire survey form

Dimension	Number	Measurement content	Reliability coefficient
Emotional investment	12	Translate interest, satisfaction, and anxiety level	0.89
Cognitive input	10	Deep thinking, strategy application, metacognition	0.86
Behavioral engagement	8	Degree of participation, persistence, and effort level	0.84
Tool experience	6	Evaluation of the usability and effectiveness of AI tools	0.91
Collaborative satisfaction	5	Satisfaction with the human-machine collaborative working mode	0.88
Translation strategy	7	Strategy selection, adjustment, and innovative application	0.82

The questionnaire data were collected through the online survey platform immediately after the translation task was completed, and participants were required to complete the questionnaire within 30 minutes, and the system set up a mandatory response function to avoid missing data. The reliability of the scale was verified by the pre-test, and the Cronbach's α coefficients were over 0.80, showing good internal consistency. The interview data were collected using a semi-structured in-depth interview method, focusing on five core themes: cognitive experience in the translation process, feelings about the use of AI tools, advantages and disadvantages of collaborative work modes, the process of adjusting translation strategies, and views on the future development of human-computer collaboration. Each interview was controlled to last 45-60 minutes, the whole process was recorded and transcribed into textual materials, and the interviewees were six representatives randomly selected from each of the two groups of participants. The whole data collection process lasted two weeks, obtaining complete and reliable research data, providing a solid foundation for the subsequent statistical analysis and theoretical construction, and enabling this study to comprehensively and deeply reveal the mechanism of human-computer synergy mode's influence on the quality of literary translation.

2.3 Data analysis

In this paper, for the collected multidimensional data, the main purpose is to carry out in-depth mining work through hybrid analysis method, combining quantitative statistical analysis and qualitative content analysis with each other, so as to explore the influence mechanism of human-computer collaborative mode on the translation quality of literary works. The quantitative analysis adopts SPSS statistical software package to comprehensively process the translation quality assessment data and questionnaire survey data, and obtains the basic distribution characteristics of each variable through descriptive statistical analysis, so as to lay the foundation for the subsequent inferential analysis. The comprehensive translation quality score is calculated using the aforementioned assessment model, and the translation quality score Q is calculated by the formula:

$$Q = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{S_i}{M_i}\right) \tag{3}$$

where n denotes the total number of reviewing experts, S_i represents the actual rating of the ith expert, and M_i is the theoretical maximum value of the expert's rating.

The significant differences between the experimental group and the control group in translation quality dimensions were compared by independent sample t-test, and Cohen's d-value was used to analyze the effect size, while analysis of variance (ANOVA) was used to test the effects of different participant characteristic variables on translation quality. In order to deeply explore the relationship between learners' inputs and translation quality, the study adopts Pearson correlation analysis to investigate the strength of the association between affective inputs, cognitive inputs, behavioral inputs and translation quality, and combines with regression analysis to design a prediction model to explore the extent to which each input dimension contributes to translation quality. The general form of the multiple regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \tag{4}$$

where Y denotes the translation quality score, X_1, X_2, X_3 are the affective, cognitive, and behavioral inputs, respectively, β_i is the regression coefficient, and ε is the random error term. Significant predictor variables were screened by stepwise regression, and the adjusted R^2 was applied to assess the goodness-of-fit of the model.

The qualitative data analysis used thematic analysis to systematically code and extract concepts from the transcribed text of the semi-structured interviews, and the analysis process followed a six-stage thematic analysis framework. The research team used NVivo 12 qualitative analysis software to assist in the coding process, with open coding to identify key concepts, axial coding to establish inter-conceptual associations, and selective coding to construct the core theoretical framework. To ensure the reliability of the coding, the initial coding was done independently by two researchers, and the divergent parts were discussed to reach a consensus. The text data of the open-ended questions of the questionnaire were used to mine the participants' deep cognition and attitudes towards the human-computer collaborative translation model through technical means such as word frequency statistics, sentiment tendency analysis, and semantic network construction. The study established a coding framework containing three main categories of positive experience, negative experience, and neutral description, with several sub-themes under each category such as technological convenience, creativity stimulation, and dependency concerns. Significant differences in the distribution of the themes among different groups of participants were analyzed by chi-square test, and correspondence analysis was used to reveal the association patterns between the groups and the themes.

3 Findings

3.1 Translation quality assessment

Based on the results of the experimental data acquisition in the previous section, this paper statistically analyzes the data under the human-computer cooperative translation mode, and its specific results are shown in Table 2.

The data show that the average score of the experimental group in the dimension of language accuracy is 8.42±0.67, which significantly exceeds the performance of the control group (6.18±0.89), and the statistical test confirms that the difference between the two groups is highly significant ($p<0.001$). This result strongly proves the important auxiliary role played

by AI tools in the basic language processing aspects such as precise vocabulary selection, grammatical structure optimization, and in-depth semantic comprehension, which significantly reduces all kinds of errors that may occur in the process of language conversion by the translators. The evaluation results of the literary retention dimension further confirm the excellent effectiveness of the human-computer synergy model, with the average score of 7.89 in the experimental group achieving a significant improvement of 37.7% compared with 5.73 in the control group, and its effect amount reaching the standard of large effect level.

The performance of the AI tool in recognizing and dealing with complex literary rhetorical techniques is surprising, especially in the understanding of difficult literary expression techniques such as metaphor, symbolism, irony, etc., which enables translators to get a more comprehensive analytical perspective and creative inspiration in the process of expression. The evaluation data of the cultural communication effect dimension also showed an encouraging positive trend, with the experimental group's scores of 8.15 and 8.26 in cultural communication effect and fluency respectively, compared with the control group's scores of 5.94 and 6.07, which is a significant improvement overall. The overall score of the experimental group is 8.18, compared with 5.98 in the control group, which is an increase of 36.8%.

Table 2: Comparison of translation quality assessment results

Dimensions	Experimental Group		Control group		Sig.
	Means	STD	Means	STD	
Language accuracy	8.42	0.67	6.18	0.89	<0.001
Maintain literary quality	7.89	0.74	5.73	0.92	<0.001
Cultural communication effect	8.15	0.58	5.94	0.85	<0.001
Overall smoothness	8.26	0.71	6.07	0.78	<0.001
Comprehensive quality score	8.18	0.62	5.98	0.81	<0.001

To summarize, artificial intelligence technology, relying on powerful information retrieval and integration capabilities, can help translators understand their cultural background more comprehensively and obtain more comprehensive translation expression strategies. So that the translator can more accurately grasp the cultural connotation of the source language and find the most appropriate expression in the cultural context of the target language. Grammar checking, style suggestion, sentence optimization and other functions that come with AI technology can help translators get high-quality translations that are more natural and fluent and in line with the expression habits of the target language.

3.2 Assessment of learner input

This study constructs an assessment framework based on the learning input theory and adopts a mixed research method combining standardized questionnaires and in-depth interviews to systematically measure and analyze the performance of the translators in the three core dimensions of emotional input, cognitive input, and behavioral input under the human-machine collaborative translation mode. The collected data were statistically analyzed to get the learners' input assessment results as shown in Table 3.

As can be seen from the table, the mean value of the experimental group in the dimension of emotional engagement is 5.89 points, which presents a significant advantage compared with the 4.23 points of the control group, and the results of the independent samples t-test show that the difference between the groups reaches a highly significant level (p<0.001), and the Cohen's d effect size is 2.15, indicating that the difference belongs to the category of large effects. Translators in the experimental group showed stronger interest-driven and

achievement satisfaction in translation tasks, and the real-time feedback mechanism and diversified suggestions provided by the AI tool effectively stimulated the translators' motivation to create, and significantly alleviated the nervousness and anxiety when facing complex literary texts. The evaluation data at the cognitive input level also confirms the positive effect of the synergistic model, with the experimental group significantly outperforming the control group with a score of 6.12 out of 4.38 out of 4.38 out of 6.12, which is 39.7% higher than the control group, and the statistical test confirms that the difference is highly significant (p<0.001). This indicates the key facilitating role played by AI assistive tools in promoting deep thinking and optimizing the use of cognitive strategies by the translators.

The positive impact of the human-computer collaboration model is further confirmed by the results of the behavioral input dimension, with the score of the experimental group significantly increased by 38.8% compared with that of the control group (p<0.001), which verifies the effect of the collaborative work approach on the initiative and continuity of the translators from the level of external behavioral manifestations. The score of the experimental group under the comprehensive input level was 5.92, which was significantly better than the 4.25 of the control group, and the Cohen's d effect size of 2.31 showed a large effect feature.

Table 3: Comparison of learner engagement assessment results

Dimension	Experimental Group		Control group		*t*	Sig.
	Means	STD	Means	STD		
Emotional investment	5.89	0.72	4.23	0.85	5.47	<0.001
Cognitive investment	6.12	0.68	4.38	0.79	6.23	<0.001
Behavioral investment	5.76	0.74	4.15	0.82	5.32	<0.001
Comprehensive level	5.92	0.65	4.25	0.78	6.08	<0.001

Table 4 shows the results of correlation and regression analysis. The correlation test revealed a strong association between learners' inputs and translation quality, with affective inputs showing a significant positive correlation with the composite score of translation quality (r=0.78, p<0.001), cognitive inputs with a correlation coefficient as high as 0.85 (p<0.001), and behavioral inputs also showing a strong positive correlation (r=0.73, p<0.001). Multiple regression analysis revealed that the three input dimensions together explained 72.4% of the variance in translation quality (R^2=0.724,F(3,20)=17.56***, p<0.001), with cognitive inputs being the most prominent predictor (β=0.52, p<0.001), affective inputs the next most important (β=0.31, p<0.001), and behavioral inputs also contributing significantly (β =0.24, p<0.001). These empirical findings fully confirm the path of human-computer collaborative model's role in improving translation quality by enhancing the learners' input level, providing important empirical support for understanding the development law of translators' competence in the age of intelligence.

Table 4: Correlation and regression analysis

Variable	Correlation		Regression		
	r	*p*	β	STD	*p*
Emotional investment	0.78	<0.001	0.31***	5.741	<0.001
Cognitive investment	0.85	<0.001	0.52***	3.682	<0.001
Behavioral investment	0.73	<0.001	0.24***	4.915	<0.001
(Con)	-	-	3.271***	4.204	<0.001
R^2	-	-	0.724		
F statistic	-	-	F (3, 20) = 17.56***, p<0.001		

3.3 Analysis of empirical data

This paper deeply integrates and analyzes the translation quality assessment data, questionnaire survey data and interview data through the triangulation validation strategy, combines the multivariate statistical method with qualitative analysis technique, and deeply explores the inner operation mechanism of the human-computer collaborative translation model and its effect characteristics. Table 5 shows the results of the comprehensive analysis of the effect of human-computer collaborative translation.

The results showed that the group factor produced a significant multivariate effect on all dimensions of translation quality (Wilks' λ=0.182, F(4,19)=21.34, p<0.001, η^2=0.818), a result that fully confirms that the human-computer collaborative model produces a substantial improvement effect on multiple dimensions of translation quality. Discriminant analysis further revealed that linguistic accuracy and cognitive engagement constituted the most important predictor variables distinguishing the two groups' translation performance, with standardized discriminant function coefficients of 0.847 and 0.723, respectively, and an overall classification accuracy as high as 95.8%, which strongly support the hypothesis of the effectiveness of the human-computer collaborative model.

Table 5: Comprehensive analysis of human-machine collaborative translation effects

Dimension	Experimental	Control	Effect size	Sig.
Translation quality	8.18±0.62	5.98±0.81	d=3.12	<0.001
Learning engagement	5.92±0.65	4.25±0.78	d=2.31	<0.001
Frequency of use	24.7±3.2	-	-	-
Number of improvements	15.3±2.8	8.6±2.1	d=2.67	<0.001
Completion time	98.4±12.6	115.7±15.3	d1.24	<0.001
Satisfaction evaluation	6.23±0.71	4.58±0.89	d=2.08	<0.001

Figure 2 shows the relationship between learning input and translation quality. As can be seen from the figure, translation quality shows a strong positive correlation with learning input (r=0.847, p<0.001). And among the three dimensions of learning input cognitive input showed the highest correlation strength (r=0.892, p<0.001), followed by affective input (r=0.756, p<0.001), and behavioral input also showed significant correlation (r=0.683, p<0.01). There is a significant positive correlation between the frequency of use of AI tools and the degree of improvement in translation quality (r=0.724, p<0.001), but this relationship is characterized by an obvious marginal diminishing effect, and the correlation strength begins to show a decreasing trend when the frequency of use exceeds the threshold of 30 times/hour. The path analysis model clearly demonstrates that the human-machine collaboration mode affects translation quality through three main paths: The direct effect path (β=0.42, p<0.001) plays a fundamental role, the learning input mediating path (β=0.35, p<0.001) serves as a bridge, and the tool-dependent moderating path (β=0.18, p<0.05) provides auxiliary support. The total effect of these three paths can explain 78.6% of the variance change.

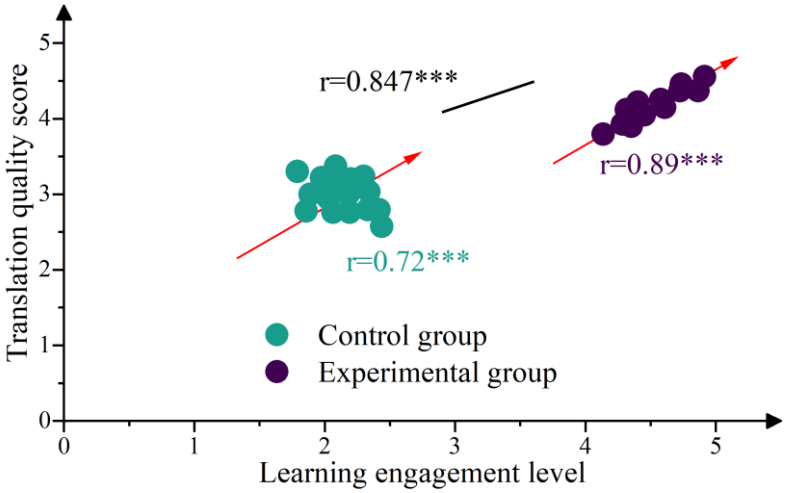

Figure 2: Learning engagement and translation quality

4 Conclusion

In this study, 24 senior undergraduate students were taken as research subjects, and the impact of artificial intelligence tools on the translation quality of literary works was analyzed based on comparative experiments. It is found that enhancing translation quality can be done by introducing the human-computer collaborative translation mode, and the students in the experimental group realize substantial improvements in the dimensions of language accuracy, literary retention, cultural communication effect, and overall fluency. Artificial intelligence can help translators better resolve the deficiencies in the translation process and improve translation quality. The findings of this paper can provide a new direction for the development of translation theory in the era of intelligence, and can also help the field of literary translation to better utilize artificial intelligence technology, which has a very clear guiding role in promoting the reform of intelligent translation education.

About the Author

Wanling He was born in Nanning, Guangxi, P.R. China, in 1988. She obtained a Master's degree from Guangxi University in China. She is currently working at the School of Foreign Studies, Guangxi Vocational Normal University. Her main research direction is translation and interpretation.

Chunlan Liang was born in Guiping, Guangxi, P.R. China, in 1984. She obtained a Master's degree from Guangxi Teachers Education University in China. She is currently working at the School of Foreign Studies, Guangxi Vocational Normal University. Her main research direction is English and American Literature.

References

[1] Dalayli, F. (2023, September). Use of NLP techniques in translation by ChatGPT: Case study. In Proceedings of the Workshop on Computational Terminology in NLP and Translation Studies (ConTeNTS) Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC) (pp. 19-25).

[2] Woodrum, C. (2024, May). ChatGPT and Language Translation: A Small Case Study Evaluating English–Mandarin Translation. In International Conference on Human-Computer Interaction (pp. 147-157). Cham: Springer Nature Switzerland.

[3] Mohammed, T. A. (2025). From Google Translate to ChatGPT: The use of large language models in translating, editing, and revising. In Role of AI in Translation and Interpretation (pp. 1-32). IGI Global Scientific Publishing.

[4] Siu, S. C. (2023). ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. Available at SSRN 4448091.

[5] Gao, R., Lin, Y., Zhao, N., & Cai, Z. G. (2024). Machine translation of Chinese classical poetry: a comparison among ChatGPT, Google Translate, and DeepL Translator. Humanities and Social Sciences Communications, 11(1), 1-10.

[6] CHENG, Y., Wang, R., Chen, J., CHAO, Y., Maimaitili, A., & Zhang, H. (2023). Context-based AI translation from a globalization perspective: A case study of ChatGPT. Sino-US English Teaching, 20(9), 370-380.

[7] Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: human translation vs. AI large language models (Gemini and ChatGPT). Cogent Social Sciences, 10(1), 2410998.

[8] Ahammad, S. H., Kalangi, R. R., Nagendram, S., Inthiyaz, S., Priya, P. P., Faragallah, O. S., ... & Rashed, A. N. Z. (2024). Improved neural machine translation using Natural Language Processing (NLP). Multimedia Tools and Applications, 83(13), 39335-39348.

[9] Li, H. (2018). Deep learning for natural language processing: advantages and challenges. National Science Review, 5(1), 24-26.

[10] Ataman, D., Birch, A., Habash, N., Federico, M., Koehn, P., & Cho, K. (2025). Machine translation in the era of large language models: a survey of historical and emerging problems. Information, 16(9), 723.

[11] Son, J., & Kim, B. (2023). Translation performance from the user's perspective of large language models and neural machine translation systems. Information, 14(10), 574.

[12] Hou, Y. (2025, July). Research on Deep Integration of Artificial Intelligence and Translation Teaching. In Proceedings of the 2025 11th International Conference on Humanities and Social Science Research (ICHSSR 2025) (Vol. 945, p. 172). Springer Nature.

[13] Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text?. In Translation quality assessment: From principles to practice (pp. 263-287). Cham: Springer international publishing.

[14] Qun, L., & Xiaojun, Z. (2014). Machine translation: general. In Routledge encyclopedia of translation technology (pp. 105-119). Routledge.

[15] Kalchbrenner, N., & Blunsom, P. (2013, October). Recurrent continuous translation models. In Proceedings of the 2013 conference on empirical methods in natural

language processing (pp. 1700-1709).

[16] Kalyani, A., & Sajja, P. S. (2015). A review of machine translation systems in india and different translation evaluation methodologies. International Journal of Computer Applications, 121(23).

[17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

[18] Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., Wong, D. F., & Chao, L. S. (2019). Learning deep transformer models for machine translation. arXiv preprint arXiv:1906.01787.

[19] Thu, T. N. T. (2020, December). Vietnamese-English Translation with Transformer and Back Translation in VLSP 2020 Machine Translation Shared Task. In Proceedings of the 7th International Workshop on Vietnamese Language and Speech Processing (pp. 64-70).

[20] Liu, H. I., & Chen, W. L. (2022). X-transformer: a machine translation model enhanced by the self-attention mechanism. Applied Sciences, 12(9), 4502.

[21] Wiesmann, E. (2019). Machine translation in the field of law: A study of the translation of Italian legal texts into German. Comparative Legilinguistics, 37(1), 117-153.

[22] Noll, R., Frischen, L. S., Boeker, M., Storf, H., & Schaaf, J. (2023). Machine translation of standardised medical terminology using natural language processing: a scoping review. New biotechnology, 77, 120-129.

[23] Matusov, E. (2019, August). The challenges of using neural machine translation for literature. In Proceedings of the qualities of literary machine translation (pp. 10-19).

[24] Skianis, K., Briand, Y., & Desgrippes, F. (2020, November). Evaluation of machine translation methods applied to medical terminologies. In Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis (pp. 59-69).

[25] Verma, S., Gupta, N., & Chauhan, R. (2022). A novel framework for ancient text translation using artificial intelligence. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 11(4), 411-425.

[26] Zhang, Y., Lou, J., & Cheng, Z. (2022). A study on the translation of cultural classics based on deep learning methods. Scientific Programming, 2022(1), 1026926.

[27] Rashed Alkatheery, E. (2023). Google translate errors in legal texts: Machine translation quality assessment. AWEJ for Translation & Literary Studies, 7(1).

[28] Vázquez, A. I. C., & Mitkov, R. (2023, September). Machine Translation of literary texts: genres, times and systems. In Proceedings of the First Workshop on NLP Tools and Resources for Translation and Interpreting Applications (pp. 48-53).

[29] Hu, K., & Li, X. (2023). The creativity and limitations of AI neural machine translation: A corpus-based study of DeepL's English-to-Chinese translation of Shakespeare's

plays. Babel, 69(4), 546-563.

[30] Wu, M., Xu, J., Yuan, Y., Haffari, G., Wan, L., Luo, W., & Zhang, K. (2025). (Perhaps) Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts. Transactions of the Association for Computational Linguistics, 13, 901-922.

[31] Karabayeva, I., & Kalizhanova, A. (2024). Evaluating machine translation of literature through rhetorical analysis. Journal of Translation and Language Studies, 5(1), 1-9.

[32] Moreno, A. I., & Mora, M. E. D. (2025). Google Translate versus DeepL in Spanish to English translation of Don Quixote. Translation and translanguaging in multilingual contexts, 11(1), 65-87.

[33] Espla-Gomis, M., Sánchez-Martínez, F., & Forcada, M. L. (2015, September). UAlacant word-level machine translation quality estimation system at WMT 2015. In Proceedings of the Tenth Workshop on Statistical Machine Translation (pp. 309-315).

[34] Fomicheva, M., Sun, S., Yankovskaya, L., Blain, F., Guzmán, F., Fishel, M., ... & Specia, L. (2020). Unsupervised quality estimation for neural machine translation. Transactions of the Association for Computational Linguistics, 8, 539-555.

[35] Rubino, R., & Sumita, E. (2020, December). Intermediate self-supervised learning for machine translation quality estimation. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 4355-4360).

[36] Lo, C. K., & Wu, D. (2013, August). MEANT at WMT 2013: A tunable, accurate yet inexpensive semantic frame based mt evaluation metric. In Proceedings of the eighth workshop on statistical machine translation (pp. 422-428).

[37] Zhou, Y., & Bollegala, D. (2019). Unsupervised evaluation of human translation quality. In KDIR: PROCEEDINGS OF THE 11TH INTERNATIONAL JOINT CONFERENCE ON KNOWLEDGE DISCOVERY, KNOWLEDGE ENGINEERING AND KNOWLEDGE MANAGEMENT-VOL 1: KDIR (Vol. 1, pp. 55-64). SCITEPRESS-Science and Technology Publications.

[38] Mukherjee, A., & Shrivastava, M. (2022, December). REUSE: REference-free UnSupervised quality Estimation Metric. In Proceedings of the Seventh Conference on Machine Translation (WMT) (pp. 564-568).