

Quality Evaluation of Technical Text Translation Based on the HACT Model

Daohan Li

School of Humanities and Education, Guangzhou Huanan Business College, Guangdong, China

Abstract: The Human-AI Collaborative Translation (HACT) model has three core modules, which are collaborative pre-editing, collaborative mid-translation, and collaborative post-editing. These modules enable dynamic, full-process collaboration between human translators and artificial intelligence systems. However, the actual effectiveness of the HACT model in the translation of technical texts has yet to be systematically tested. To address this gap, this study employs DeepSeek as an empirical tool, selects a corpus of technical texts, and introduces the Multidimensional Quality Metrics (MQM) framework for error classification and quality assessment. By comparing the quality differences between raw machine translation outputs and HACT-mediated collaborative translations across four dimensions—terminology errors, logical errors, accuracy, and fluency—this study examines the extent to which the HACT model improves translation quality.

Keywords: HACT Model; Multidimensional Quality Metrics (MQM); Technical Texts; Translation Quality Assessment; DeepSeek

1. Introduction

According to Wang and Zhang (2026), Human-AI Collaborative Translation (HACT) refers to a translation model that integrates translators' professional expertise and judgment to achieve intelligent collaboration across the entire process, from pre-translation to translation and post-editing. Wang and Zhang (2025) further identified that this model consists of three core modules: collaborative pre-editing, collaborative mid-translation, and collaborative post-editing. The core of this process is not translation but strategic decision-making: the translator transforms their cross-cultural discourse knowledge into editing instructions, while the model executes these instructions. To systematically validate the practical utility of the

HACT model in translating technical texts, this study adopts the Multidimensional Quality Metrics (MQM) framework as the basis for classifying and grading translation errors. This allows each instance of translator feedback and system adjustment generated during the HACT collaborative process to be systematically identified, categorised, and quantitatively evaluated. Using DeepSeek as the translation tool. This study selects excerpts from Problem Solving Activities in Post-editing and Translation from Scratch as the corpus. This study compares the MQM penalty scores between raw DeepSeek outputs and HACT outputs across four dimensions: terminology, logic, accuracy, and fluency. It then investigates whether the HACT model improves technical translation quality, providing empirical evidence for promoting and optimising HACT in technical translation.

2. Theoretical Foundations for Evaluating Translation Quality in Technical Texts under the HACT Model

2.1 Linguistic Features of Technical Texts and Core Quality Requirements

Technical texts are highly specialized, objective, and information-dense. Their linguistic features include dense terminology, complex syntax, rigorous logic, and standardized style. Passive voice and nominalization are common, making the style objective rather than subjective. Also, the frequent use of specialized terminology places high demands on translators' domain knowledge. As Liu, Yang, and Wang (2023) noted, when translating lexical items in technical texts, translators should focus on whether the translated terms provide sufficient contextual effects and optimal relevance for the reader. Based on these features, the core quality requirements for technical text translation can be summarized as terminology, accuracy, logic and fluency.

2.2 the Structural Modules and Operational Mechanisms of the HACT Model

Neural machine translation and large language models have developed rapidly. As a result, AI's role in translation has shifted from an initial assistive tool to an intelligent agent. This agent can now collaborate deeply with human translators. However, most current human-AI collaborative models still operate at a shallow level. They mainly follow a "translate-and-revise" pattern, failing to fully use the cognitive strengths of both humans and machines. Against this backdrop, the HACT model has emerged as an advanced form of human-AI collaborative translation. The HACT model refers to a new collaborative approach. In this approach, human translators and AI systems both engage in dynamic, continuous interaction and feedback. This enables deep cooperation in translation generation, refinement, and critical decision-making. Guided by the concept of human-machine symbiosis, the HACT model consists of three modules: collaborative pre-editing, collaborative mid-translation, and collaborative post-editing. Based on this positioning, the model further operationalises a framework built on these three modules. Liu and Zhang (2025) found that the pre-editing module requires translators to analyse and process the input text and then issue editing instructions to the model. The mid-translation module emphasises real-time interaction between the translator and the model. Translators design prompt instructions to guide the generation of personalised translations. The post-editing module focuses on detailed revision. It advances optimisation through a "revision-feedback-re-revision" cycle. These three interdependent modules together form a complete translation chain.

2.3 The Classification System and Grading Standards of the MQM Framework

The Multidimensional Quality Metrics (MQM) framework is currently one of the most standardised error classification and assessment systems in the translation industry. In terms of grading standards, MQM categorises errors into three levels-Critical, Major, and Minor-based on the degree to which they affect the target reader's understanding and use of the text. Different levels are assigned different penalty weights, providing a quantitative basis for final quality scores. This grading mechanism allows the

severity of errors to be objectively quantified, thereby effectively distinguishing critical defects that affect overall comprehension from minor flaws that do not compromise overall quality. It thus provides an actionable standard for the refined management of translation quality (Multidimensional Quality Metrics Community, n.d.).

The classification dimensions of the MQM framework cover accuracy, fluency, terminology, style, and other aspects, with each dimension being both independent and systematically interrelated. For the specific assessment needs of technical texts, the weighting of the accuracy and terminology dimensions is particularly critical, because deviations in terminology in technical texts often lead to logical comprehension difficulties. Measuring such texts solely by fluency indicators would distort the assessment results. MQM allows evaluators to customise the weighting of each dimension according to text type and use scenario, which gives it strong adaptability across different domains. Importantly, MQM is not a static scoring tool but a dynamic quality assessment system. By using it, researchers can build a consensus on evaluation, thereby effectively reducing the influence of subjective judgments in this scenarios. This feature is particularly evident in dimensional discussions and provides methodological assurance for this study.

2.4 Adaptability Analysis of Integrating HACT and MQM

The HACT model and the MQM framework are highly complementary in their logical structures. The HACT model generates translatable outputs through a three-stage human-AI collaborative process-pre-editing, mid-translation, and post-editing. While the MQM framework performs a fine-grained function in identifying and defining specific errors, providing a unified error classification language and quantitative basis for quality judgments across dimensions. Their integration creates an effective link between the macroscopic collaborative process and the microscopic error annotation, enabling evaluation results to have both cross-module comparative value and longitudinal error-tracing capability. Technical text translation has specific characteristics. The HACT model focuses on terminological consistency and semantic precision during the mid-translation and post-editing stages. This focus resonates

intrinsically with MQM's prioritisation of terminology and accuracy errors. Therefore, the combined approach offers significant advantages when assessing specialised texts.

3. A Quality Assessment Scheme for Technical Text Translation Based on the HACT Model

3.1 Corpus Selection Principles and Preprocessing for the DeepSeek-Based Empirical Study

This study selects as its analytical object technical text corpora processed through the HACT model using DeepSeek (hereinafter referred to as DS) as a translation tool. The corpus is taken from Problem Solving Activities in Post-editing and Translation from Scratch by Jean Nitzke, which systematically explores behavioural differences and cognitive loads in post-editing versus translation from scratch, covering experimental design, data analysis, and specialised terminology. This text exhibits a high density of terminology and also has rigorous logical structures. In addition, it contains many passive constructions and long sentences. These features meet this study's definition of technical texts. In the preprocessing stage, the corpus underwent text cleaning, sentence segmentation, and source-target alignment. This ensured accurate sentence-by-sentence correspondence. A domain-specific term list was also established to provide a reference baseline for subsequent error annotation, ensuring that empirical data collection could proceed systematically under standardised conditions.

3.2 Operational Procedures and Prompt Design for Each HACT Collaborative Module

To provide a directly operable practical workflow for the three core dimensions of the HACT model, this study specifies each module for technical texts. In the pre-editing stage, the translator identifies logical and referential issues in the source text and issues editing instructions to reorganise the information structure. In the mid-translation stage, the translator designs customised prompts and uses examples to guide the model's output. In the post-editing stage, multiple rounds of revision focus on terminology and technical expressions. To facilitate subsequent comparative quality analysis, this study also retains the raw DS outputs as a control, allowing systematic evaluation of the

actual improvement in technical text translation quality resulting from the full HACT intervention.

3.3 Error Annotation Procedure for HACT-Mediated Translations Based on MQM Indicators

Following the MQM error classification system, this study designs a layered annotation procedure. The procedure is divided into three stages: identification, classification, and grading. This ensures that each stage has clear operational objectives and remains independent of the others. In the identification stage, annotators compare the raw DS outputs and the HACT-mediated translations with the source text sentence by sentence. In the classification stage, we refer to the MQM framework and the characteristics of technical texts. We also analyze the raw outputs. Based on these, four evaluation dimensions are selected. They are terminology errors, logic, fluency, and accuracy. In the grading stage, each identified error is assigned a score according to a four-level severity scale: neutral (0 points), minor (1 point), major (5 points), and critical (10 points), based on the error's impact on reader comprehension and information integrity. The entire annotation process is recorded in a standardised annotation table to ensure uniform data formatting. A combination of group discussion and expert consultation is used to reconcile scores. Any disagreements are going to be discussed again in order to ensure the reliability and validity of the annotation results.

3.4 Data Collection, Organisation, and Reliability Assurance Methods

Data collection is conducted using a structured annotation table. The table records scores and error details for each sentence segment, ensuring that the data granularity meets the requirements for subsequent statistical analysis. In the data organisation stage, raw annotations are aggregated and summarised. The score distributions for HACT processed texts, along with the frequency and severity composition of each error category, are then calculated separately. These calculations lay the empirical foundation for the experimental results. For reliability assurance, this study employs a combination of group discussion and expert consultation, conducting multiple rounds of review of the annotation results. All scoring disagreements are resolved through

second-round discussions and final confirmation by experts to ensure that the evaluation conclusions are not systematically biased by individual annotator differences. In addition, cross-validation is performed after data entry to eliminate missing or duplicate annotation items, ensuring the completeness and accuracy of the dataset. These data management measures collectively constitute the quality control mechanism of the empirical scheme, enabling subsequent analysis and validation to be built on a reliable and stable data foundation.

Table 1. MQM Penalty Scores for Raw DS Outputs

	Error type	Neutral (0)	Minor (1)	Major (5)	Critical (10)	Error type weight	Penalty type total
1	Terminology	0	8	2	0	0.25	18
2	Logic	0	7	2	0	0.25	17
3	Fluency	0	0	2	0	0.25	10
4	Accuracy	0	4	12	1	0.25	64
					Absolute total penalty:		109
Number of words evaluated	1003				Per-Word Penalty Total		0.259
Reference word count	1003				Total Normed Penalty Points		0.284
Scaling Parameter(SP)	25				Total Quality Score		72.85
Maximum Score Passing Threshold	100						
Passing Threshold	75				Pass / Fail		Fail

Table 2. MQM Penalty Scores for HACT Model Outputs

	Error type	Neutral (0)	Minor (1)	Major (5)	Critical (10)	Error type weight	Penalty type total
1	Terminology	0	0	0	0	0.25	0
2	Logic	0	0	3	0	0.25	15
3	Fluency	0	1	1	0	0.25	6
4	Accuracy	0	2	1	0	0.25	7
					Absolute total penalty:		28
Number of words evaluated	914				Per-Word Penalty Total		0.0306
Reference word count	914				Total Normed Penalty Points		0.044
Scaling Parameter(SP)	25				Total Quality Score		93
Maximum Score Passing Threshold	100						
Passing Threshold	75				Pass / Fail		Pass

The comparison between Table 1 and Table 2 clearly illustrates the differences in penalty scores between the raw DS outputs and the HACT-mediated translations across the four error dimensions: terminology, logic, fluency, and accuracy.

In terms of penalty distribution, the raw DS outputs show that accuracy errors dominate, with a single-dimension penalty of 64 points, accounting for 58.7% of the total penalty (109). The penalties are concentrated at the Major and Critical levels, making accuracy the core factor lowering translation quality. This is followed by terminology errors (18 points, 16.5%) and logical errors (17 points, 15.6%). Fluency errors

4. Translation Quality Analysis

4.1 Error Types and Distribution Characteristics of HACT-Mediated Translations Based on MQM

Based on the error annotation results obtained from the MQM framework, this study systematically examined the error types and distribution characteristics of the raw DeepSeek (DS) outputs and the HACT outputs (see Tables 1 and 2).

are the lowest, at only 10 points, and are not the primary cause of quality loss. In contrast, the HACT-mediated translations show a substantial reduction in total penalty, from 109 to 28 points. Logical errors remain relatively high (15 points), but fluency and accuracy errors are reduced to only a few minor and a small number of major errors, with no critical errors. Terminology errors are completely eliminated. Overall, the quality of the HACT-mediated translations is significantly improved. As shown by the penalty comparison in Tables 1 and 2, the HACT collaborative process markedly enhances translation quality. The raw DS outputs

received a total quality score of 72.85, falling below the passing threshold of 75. In contrast, the HACT-mediated translations achieved a score of 93 which shows a substantial improvement that comfortably meets the passing standard. In terms of error distribution, accuracy errors in the raw DS outputs account for 58.7% of the total penalty (64 out of 109 points), making them the single largest penalty contributor. When combined with terminology and logical errors, these three categories account for over 90% of the total penalty, with most penalties concentrated at the Major and Critical levels. By comparison, the HACT outputs achieved zero penalty for terminology errors, with no neutral, minor, or major errors. Only logical errors retain a relatively high penalty (15 points). However fluency and accuracy errors are reduced to a few minor and a small number of major errors, with no critical errors. These results indicate that the HACT model improves technical text translation primarily through three mechanisms: standardised control of terminology, substantial reduction of high-severity errors, and optimisation of translation accuracy.

4.2 Influencing Factors of HACT Outputs

Four main factors influence the translation quality of the HACT model. First, the quality of the researcher's prompt instructions is critical. The more specific and actionable the instructions, the more reliable the model's output. For instance, in correcting a logical error, the researcher explicitly instructed the model to "make the conditional logic explicit by using 'therefore' ", and the model generated a logically clear translation accordingly. According to Wen and Liang (2024), human-AI interaction negotiation competence encompasses five key elements: understanding the AI, setting goals, issuing instructions, analysing feedback, and adjusting strategies. The refined design of prompts is a core prerequisite for improving collaborative translation effectiveness.

Second, the DeepSeek model has limitations in comprehending the implicit contextual semantics of the source text. Wang and Zhang (2024) noted that large language models commonly suffer from content hallucinations and partial contextual understanding, making them unable to fully discern implicit meanings independently. Consequently, translators must continuously supply contextual information to complete the

revision process. The model often quickly adopts user suggestions but may only make superficial changes. For example, in a fluency correction, the model's initial output "staring at" was more natural than a literal translation but still lacked academic appropriateness. The researcher explained to DeepSeek that the process of "盯着看" actually involved repeatedly reconsidering whether the translation was correct; after this guidance, the output was refined to "反复斟酌". Third, terminological standardisation cannot rely entirely on the model, especially when academic controversies are involved. In such cases, expert consultation becomes necessary. For example, the translation of "gaze data" was determined to be "注视数据" only after expert consultation, rather than the model's output "视线数据".

Fourth, the number of interaction rounds and the depth of negotiation play a crucial role. Multiple "revision-feedback-re-revision" cycles are particularly important for complex sentence structures and context-dependent texts. The four correction cases analysed in this study required an average of 2-4 interaction rounds, with each round showing improvement over the previous version.

Together, these four factors determine the actual effectiveness of the HACT model and provide a reference for optimising the collaborative workflow in future applications.

Conclusion

The integrated application of the HACT model and the MQM multidimensional quality metrics framework provides a structured, quantifiable evaluation system for technical text translation quality. Using DeepSeek as the collaborative tool, the HACT practice confirms that the "revision-feedback-re-revision" mechanism effectively drives overall translation improvement. At the same time, the MQM error classification and grading system offers a standardised basis for quality assessment.

The study further identifies four factors that influence HACT output quality. One critical factor is prompt quality: specific and actionable instructions produce more reliable model outputs. Another limitation lies with DeepSeek itself, as it struggles to understand implicit contextual semantics; therefore, translators must continuously supply contextual cues to complete revisions. A further issue concerns terminological standardisation, which cannot rely solely on the model-especially when

academic controversies exist, making expert consultation necessary. The last factor is the number of interaction rounds and the depth of negotiation, as multiple "revision-feedback-re-revision" cycles prove essential for complex or context-dependent sentences. Together, these four factors determine the actual effectiveness of the HACT model and guide future workflow optimisation.

This evaluation system offers a rigorous pathway for scientifically judging technical text translation quality. By combining the HACT collaborative framework with the MQM quantitative assessment system, iterative optimisation of translation teaching and professional practice can be driven toward greater standardisation and intelligence, with positive implications for advancing technical text translation in a standardised and professional direction.

References

- [1]Liu, C., Yang, C., & Wang, T. (2023). An overview of scientific text translation research from the perspective of relevance theory. *Language and Culture Research*, 26(1), 123-125.
- [2]Wen, Q., & Liang, M. (2024). Human-computer interaction negotiation competence: ChatGPT and foreign language education. *Foreign Language Teaching and Research*, 56(2), 286-296.
- [3]Liu, S., & Zhang, Y. (2025). From MTPE to HACT: A study on innovation in translation process driven by large language models. *Foreign Language Education Research*, 13(1), 18-26.
- [4]Wang, Y., & Zhang, Z. (2024). Concerns and solutions of ChatGPT AI translation. *Chinese Translators Journal*, (2), 95-102.
- [5]Multidimensional Quality Metrics Community. (n.d.). Multidimensional Quality Metrics (MQM). Retrieved June 5, 2026, from <https://themqm.org/>
- [6]Wang, J., & Zhang, Z. (2026). Construction of human-AI collaborative translation model from the perspective of human-machine symbiosis. *Shanghai Journal of Translators*, 1(1), 13-19.
- [7]Wang, J., & Zhang, Z. (2025). A large language model-driven human-machine collaborative translation teaching model. *Foreign Language World*, (5), 69-76.