Optimizing Language Service Efficiency Through Machine Translation and Human Proofreading Collaboration in Cross-Cultural Communication Scenarios

Zhaoxia Liu[1], Yingying Du[2], Xiaohui Zhang[3]
[1]School of Foreign Languages, University of Sanya, Sanya 572022, Hainan, China
[2]Institute for Development and Security Studies, Jilin University, Panshi 132311, Jilin, China
[3]Hainan College of Foreign Studies, Wenchang 571300, Hainan, China
Corresponding author: Xiaohui Zhang (e-mail: Jlsfdx007@163.com)

ABSTRACT To address the conflict between machine translation and human proofreading in cross-cultural communication, this paper integrates language-service quality assessment with collaborative optimization theory and establishes an MT-HC collaborative efficiency optimization framework. Machine translation has advantages in speed but suffers from unstable accuracy and insufficient cultural adaptation, while human proofreading ensures quality but is costly and inefficient. The study defines three application scenarios: business negotiation, academic exchange, and public service, and specifies accuracy, cultural-adaptation, and time constraints for each. A three-stage workflow of "machine translation pretranslation-intelligent error classification-proofreading priority ranking" is designed. The BERT model, with an error recognition F1-score of 0.92, and fuzzy analytic hierarchy process are integrated to construct a proofreading-priority matrix. Comparative experiments show that the optimized model reduces business translation cycles by 42%, lowers academic proofreading costs by 38%, and decreases cultural-adaptation error rates in public services by 51%. It increases effective translation volume per unit time by 2.3 times and improves efficiency in minor-language scenarios by at least 1 .8 times.

INDEX TERMS cross-cultural communication, machine translation, human proofreading, collaborative optimization, error classification

I. INTRODUCTION

A. BACKGROUND AND SIGNIFICANCE OF THE STUDY

AGAINST the backdrop of increasingly frequent global exchanges, cross-cultural communication imposes three critical demands on language services: timeliness, precision, and cultural adaptability, especially when translating intricate material specifications and branding content for the international market.

Machine translation technologies (such as Google Translate and Baidu Translate), leveraging their advantages of high speed and broad coverage, have become foundational tools for cross-cultural communication.

However, they still exhibit significant shortcomings in handling specialized terminology (e.g., "force majeure" in business contracts), cultural metaphors (e.g., the semantic difference between Chinese "龙" and English "dragon"), and complex sentence structures (e.g., lengthy, intricate sentences in academic papers). This results in translation accuracy fluctuating between 55% and 85%, making it difficult to meet the demands of high-stakes scenarios [1].

Human proofreading, serving as the "final line of defense" for MT quality, can correct semantic deviations and cultural errors but faces two major bottlenecks: First, an efficiency bottleneck—under the traditional "sentence-by-sentence proofreading" model, professional translators process only 8,000–12,000 words daily, far below the million-word scale of MT. Second, cost constraints: manual proofreading for less-common languages can cost $0.15–0.30 per word, making sustained use in high-frequency cross-cultural communication scenarios (e.g., cross-border e-commerce customer service, simultaneous interpretation support at international conferences) financially unsustainable [2]. The core conflict in their collaboration lies in the failure to complement MT's speed advantage with HC's quality advantage—overreliance on MT risks quality issues, while overreliance on HC leads to inefficiency.

From a practical perspective, this study's optimization framework can be directly applied to three typical cross-

cultural scenarios: In business negotiations, "MT rapid draft generation + HC focused core clause proofreading" reduces translation response time from 24 hours to 4 hours; In academic exchanges, "intelligent error classification" reduces proofreading effort for non-critical errors by 80%; In public service scenarios (e.g., embassy visa document translation), a "cultural adaptation rule repository" reduces cultural misunderstanding risks by 60% [3]. Theoretically, this research breaks from the traditional "linear collaboration between MT and HC" paradigm by establishing a dynamic matching model linking "error types- proofreading resources-scenario requirements." This expands the research dimensions of collaborative optimization in language services and provides new pathways for implementing natural language processing technologies in cross-cultural communication, including the critical task of accurately translating specialized terminology for advanced manufacturing and international standards.

B. CURRENT STATUS OF RESEARCH

Machine translation research has evolved from rule-based and statistical approaches to the current stage of neural machine translation (NMT). However, significant shortcomings persist in cross-cultural contexts: In 2021, Li et al. compared the performance of 10 mainstream MT systems in translating business contracts, finding t hat t he average accuracy rate for specialized terminology was only 68%, with errors in expressions related to "legal liability" reaching as high as 32% [4]. In 2023, Wang's team demonstrated in academic paper abstract translation studies that NMT achieved less than 70% accuracy in handling passive voice and logical connectives, leading to distortions in conveying scholarly arguments [5]. In 2024, Zhang et al. pointed out that MT systems exhibit over 45% error rates in translating cultural metaphors. For instance, the Chinese idiom "拱手相让" (literally "hand over with hands folded") was mistranslated as "handover with hands folded," failing to convey its core meaning of "voluntarily relinquishing rights" [6]. Existing research predominantly focuses on optimizing MT technology itself, lacking collaborative design with human proofreading.

Research on optimizing human proofreading efficiency primarily revolves around "process simplification" and "tool assistance": In 2019, Chen et al. proposed the "Keyword-Priority Proofreading Method," which extracts professional terminology and culturally sensitive terms from MT texts for prioritized review, boosting efficiency b y 1 5% b ut failing to achieve intelligent error-type recognition [7]. In 2022, Xu's team developed an "MT Error Annotation Tool" capable of automatically flagging g rammatical e rrors, yet its recognition rate for semantic deviations and cultural adaptation errors remained only 52% [8]. In 2023, Liu et al. introduced a "Dynamic Resource Allocation for Proofreading" mechanism, assigning proofreading tasks for different error types based on translators' expertise, achieving a 23% efficiency g ain, b ut f ailing t o a ddress t he v arying demands across cross-cultural contexts [9]. Existing research has not established a complete collaborative chain encompassing "MT error prediction → proofreading resource matching → scenario requirement adaptation." Current MT-HC collaborative research suffers from two shortcomings: First, the collaborative logic is overly simplistic. Most studies employ a linear workflow of "MT generates a draft → HC revises sentence by sentence," failing to prioritize proofreading based on error severity. This results in 80% of proofreading resources being wasted on non-critical errors (e.g., punctuation, simple syntax) [10]. Second, insufficient scenario adaptation fails to account for differing demands across cultural communication contexts—where business scenarios prioritize terminology accuracy, academic scenarios emphasize logical coherence, and public service scenarios demand cultural adaptation—resulting in non-targeted collaborative strategies [11].

II. CORE THEORETICAL AND TECHNICAL FOUNDATIONS

A. DIFFERENCES IN LANGUAGE SERVICE REQUIREMENTS ACROSS CROSS-CULTURAL COMMUNICATION SCENARIOS

Based on the "purpose, audience, and risk cost" framework of cross-cultural communication, language service scenarios are categorized into three types: business negotiations, academic exchanges, and public services. Each scenario exhibits distinct core requirements and constraint standards. Business Negotiation Scenario: The purpose is to ensure precise communication of contract terms and cooperation proposals. The audience consists of corporate decision-makers and legal professionals, with high risk-cost (translation errors may result in millions in economic losses). Core requirements include "technical term accuracy $\geq$95%, unambiguous legal phrasing, turnaround time $\leq$8 hours." Typical text types include contract drafts, quotations, and memoranda. Academic exchange scenarios aim to facilitate cross-border dissemination of research findings, targeting researchers. These scenarios carry moderate risk costs (misinterpretations may impact academic evaluations). Core requirements include: "logical coherence $\geq$90%, academic terminology consistency $\geq$92%, turnaround time $\leq$72 hours." Typical text types include abstracts, conference presentations, and research reports. Public service scenarios aim to resolve daily communication needs, targeting the general public. These scenarios carry low risk costs but broad impact (cultural misunderstandings may trigger public relations issues). Core requirements include "cultural adaptability $\geq$85%, colloquialism level tailored to target audience, turnaround time $\leq$24 hours." Typical text types include visa documents, public notices, and customer service dialogues [12-16], as shown in Table 1 below.

TABLE 1. Language Service Constraint Standards for Cross-Cultural Communication Scenarios

Scenario Type	Cultural Adaptation Constraint	Accuracy Constraint (Key Information)	Timeliness Constraint	Cost Constraint (USD per 1,000 Words)
Business Negotiation	No cultural metaphor conflicts, in line with target country's business practices	≥95%	≤8 hours	≤50
Academic Exchange	Standard academic expression, no logical ambiguity	≥90%	≤72 hours	≤30
Public Service	In line with target country's cultural customs (e.g., avoidance of taboo terms)	≥85%	≤24 hours	≤15

B. ERROR TYPES AND DETECTION TECHNIQUES IN MACHINE TRANSLATION

This study employs a "pre-trained language model (BERT) + multi-label classification" architecture to achieve automatic identification of MT error types. The specific technical approach is as follows: First, MT texts and manually proofread correction data from three scenarios (totaling 100,000 sentences, each annotated with error type) were collected to construct the annotated dataset; Second, using BERT-Base as the base model, we added "scenario type embeddings (e.g., business/academic/public service)" to the input layer and employed a sigmoid activation function in the output layer to enable simultaneous recognition of multiple error types. Subsequently, we enhanced the recognition accuracy of error types with small samples through "dynamic learning rate adjustment (initial 0.001, decaying by 0.9 every 10 epochs)" and category weight balancing (doubling weights for culturally conflicted samples due to scarcity)", we enhanced recognition accuracy for low-frequency error types. Ultimately, on the test set (20,000 sentences), the model achieved an overall F1 score of 0.92, with F1 scores $\geq$0.95 for terminology errors and semantic deviations, and $\geq$0.88 for cultural conflicts, meeting the error screening requirements for collaborative optimization [17].

C. RESOURCE ALLOCATION AND EFFICIENCY OPTIMIZATION THEORY FOR HUMAN PROOFREADING

Human proofreading resources encompass three categories—"translator expertise, time investment, and tool support"—whose alignment with scenario requirements directly impacts collaborative efficiency.

Translator expertise must align with the scenario: business contexts require "legal/financial background + bilingual terminology databases"; academic contexts demand "disciplinary background + scholarly writing experience"; public service contexts necessitate "target-country cultural knowledge + colloquial expression skills." Time allocation should reflect error-type correction difficulty—e.g., terminology errors average 15 seconds/correction, while cultural conflict errors average 60 seconds/correction (requiring contextual adaptation). Tool support includes:

- Real-time terminology matching (e.g., auto-suggesting "force majeure" when inputting 'force majeure') - Cultural error alerts (e.g., flagging "dragon" for its Western cultural connotations) - Revision traceability (facilitating subsequent MT model fine-tuning) [18-22].

To avoid data leakage and overfitting risks during monthly fine-tuning:

Data Leakage Prevention: Time-series partitioning is adopted instead of random division for training and validation sets, ensuring validation set data consists of newly added texts after the training set period. Data isolation technology is implemented to shield validation set features during training, which is only used for performance evaluation without participating in parameter updates.

Overfitting Control Strategies:

1. L2 regularization with weight decay coefficient λ=0.001 to limit parameter scale;
2. Early stopping mechanism (training stops if validation set accuracy improvement is <0.5% for three consecutive rounds);
3. Strict control of fine-tuning data volume, using only 5,000 high-quality correction samples added monthly to avoid overfitting to specific error patterns.

III. DESIGN OF THE MT-HC COLLABORATIVE EFFICIENCY OPTIMIZATION FRAMEWORK

A. OVERALL FRAMEWORK ARCHITECTURE AND KEY TECHNICAL MODULES

This study constructs a "fourth-order dynamic collaborative framework" to achieve deep integration of "MT speed advantages" and "HC quality advantages." The framework process sequentially follows: "MT pre- translation $\rightarrow$ intelligent error classification $\rightarrow$ human proofreading priority ranking $\rightarrow$ quality feedback iteration." During the MT pre-translation phase, an appropriate MT model is selected based on the scenario type (e.g., "legal domain NMT model" for business scenarios, "subject-customized model" for academic scenarios). The source text is input to generate a draft, simultaneously outputting foundational metrics such as "terminology matching rate" and "sentence complexity" [23-25]. The Intelligent Error Classification stage feeds the MT draft into a BERT-based error detection model. By integrating scenario-specific embeddings, it outputs error type labels (e.g., "terminology error + semantic deviation") and confidence scores (e.g., 0.95) for each sentence, filtering out low-risk errors with confidence <0.6. Manual Proofreading Priority Ranking Phase: Using the FAHP evaluation matrix, calculates priority scores for each error based on "error severity, correction cost, and scenario sensitivity." Generates proofreading task lists categorized as "High Priority (Score $\geq$ 0.8), Medium Priority (0.5 $\leq$ Score < 0.8), Low

Priority (Score < 0.5)." High-priority errors are assigned to senior translators first. The quality feedback iteration phase collects correction data from manual proofreading (error type, correction content, scenario label). This data updates the "Cross-Cultural Error Type Library" while also feeding high-frequency error data into MT model fine-tuning (e.g., monthly fine-tuning to update model parameters), thereby enhancing the quality of subsequent MT drafts [26].

Error Label Quantitative Conversion Module:To connect BERT-based qualitative error labels with FAHP-based quantitative priority scores, a three-step conversion process is designed:

BERT outputs error type labels (e.g., 'Professional Term Deviation') and confidence scores (0-1);

FAHP determines the severity weight of each error type (e.g., 0.3 for legal term deviations and 0.05 for punctuation errors in business scenarios);

Quantitative score = Error confidence $\times$ Type weight $\times$ Scenario sensitivity coefficient (0-1, dynamically adjusted by scenario), realizing accurate conversion from qualitative labels to quantitative scores.

Example: For 'legal term deviation in business scenarios' - BERT recognition confidence = 0.95, FAHP severity weight = 0.3, scenario sensitivity coefficient = 1.0. Quantitative score = 0.95 $\times$ 0.3 $\times$ 1.0 = 0.285, which is used for priority ranking in the FAHP matrix.

Key technical modules include the Cross-Cultural Error Type Library, Proofreading Priority Assessment Model, and Quality Feedback Iteration Mechanism. The Cross-Cultural Error Type Library is constructed based on practical requirements across three scenario categories, covering "6 major categories and 23 subcategories" of errors. It clearly defines each error type's "definition, identification characteristics, and correction principles." For example, "terminology deviation" under terminology errors refers to inconsistencies between professional field terminology translations and standard expressions. Its identification characteristic is the presence of specific vocabulary from "legal/academic/public service" domains with a termbase match rate <80%. The correction principle involves consulting scenario-specific terminology databases (e.g., "Incoterms" for business contexts). "Metaphorical semantic conflicts" in cultural clashes refer to semantic differences in cultural symbols (e.g., animals, colors). Identification features include the presence of culturally sensitive symbols like "dragon, phoenix, white" carrying negative connotations in the target language context. The correction principle involves replacing them with culturally neutral symbols or adding explanatory notes [27].

The priority assessment model for proofreading is based on the Fuzzy Analytic Hierarchy Process (FAHP), constructing a dynamically weighted priority evaluation framework. The core formula is $P_i = w_s S_i + w_c(1-C_i) + w_T T_i$, where P_i represents the priority score of an error at position i (0–1), and w_s, w_c, w_T denote the weighting factors for "error severity (S), correction cost (C), and scenario sensitivity (T)" (satisfying $w_s + w_c + w_T = 1$, dynamically adjusted per scenario). S_i denotes the error severity score (0–1, where high-risk errors score 0.8–1.0 and low-risk errors score 0.2–0.5). C_i represents the correction cost score (0–1, correction time < 30 seconds scores 0.2–0.5, correction time > 60 seconds scores 0.8–1.0, $(1 - C_i)$ indicates higher priority with lower cost), T_i is the scenario sensitivity score (0–1, errors related to core scenario requirements score 0.8–1.0, non-core requirement errors score 0.3–0.6) [28–30].

The quality feedback iteration mechanism establishes a closed-loop process of "human proofreading $\rightarrow$ data cleansing $\rightarrow$ model fine-tuning $\rightarrow$ effectiveness validation." The specific workflow is as follows: Collect correction data containing "source text, MT draft, HC-corrected draft, error type labels, and scenario labels";

Filter out non-substantive error data such as "typo corrections and formatting adjustments," retaining core data like "semantic corrections and cultural adaptation adjustments"; Employing "incremental training," the cleaned data is divided into a 1:4 ratio for validation and training sets. The MT model undergoes monthly fine- tuning (with a learning rate set to 1/10 of the initial training to prevent model forgetting). Post-fine-tuning, the accuracy improvement of the MT draft on the test set is verified. If the accuracy improvement reaches $\geq$5%, the online model is updated; otherwise, the reasons are analyzed and additional data is supplemented [31].

Professional Term Deviation Recognition Process:

Preliminary screening: Identify terms with abnormal matching rates (<85% or 100%) based on the term library;

Semantic verification: Extract a 10-word contextual window (5 words before and after the term) and calculate semantic fit using the pre-trained BERT model;

Scenario adaptation: Exclude scenario-inappropriate terms (e.g., legal terms in medical contexts) to confirm deviations.

B. SCENARIO-BASED COLLABORATIVE STRATEGY DESIGN

Differentiated collaborative strategies must be designed to address the distinct needs across three types of cross-cultural scenarios. For business negotiation scenarios, adopt a "high precision-fast response" strategy:

MT pre-translation utilizes a "legal domain NMT model," automatically loading the "International Trade Terminology Database" and "Legal Expression Rules Database" when inputting source text (e.g., contract drafts). Error classification prioritizes identifying three high-risk error types—"terminology errors, semantic deviations, and ambiguous legal expressions"—while filtering out "punctuation errors and minor grammatical errors" (automatically corrected by tools). Prioritizing errors: High-priority errors (e.g., incorrect wording in contract clauses like "payment methods" or "breach of contract liabilities") are assigned to senior legal translators; medium-priority errors (e.g., standard business terminology translation) go to regular translators; low-priority errors (e.g., formatting issues) are automatically

handled by the system. Timeliness is ensured through "batch proofreading": core clauses (e.g., pricing, delivery timelines) are proofread within 1 hour; non-core clauses (e.g., appendix notes) within 4 hours; overall turnaround $\leq$8 hours [32-35], as shown in Table 2 below.

TABLE 2. Cross-Cultural Error Type Library

Error Category	Error Subcategory	Definition	Recognition Features	Correction Principles
Term Error	Professional Term Deviation	Inconsistency between the translation of a professional term and its standard expression	Appearance of domain-specific vocabulary in "legal/academic/public service" fields, with a term library matching degree < 80%	Correct by referring to scenario-specific term libraries (e.g., "Incoterms" for business scenarios)
Term Error	Term Consistency Error	Multiple translations of the same term in the same text	The same vocabulary appears $\geq$2 times in the text, with a translation difference rate > 10%	Unify into the recommended translation for the scenario to ensure consistency throughout the text
Cultural Conflict	Metaphorical Semantic Conflict	Semantic differences of cultural symbols (such as animals, colors)	Appearance of culturally sensitive symbols like "dragon, phoenix, white", with negative connotations in the target language context	Replace with neutral symbols in the target culture or supplement with explanatory notes
Cultural Conflict	Custom Adaptation Error	Expression that does not conform to the etiquette and customs of the target country	Appearance of "addresses, greetings, time expressions" with a matching degree < 70% to the target country's customs	Adjust the expression by referring to the target country's cultural customs guide (e.g., avoid "Dear Sir/Madam" in English emails and use specific titles instead)

For academic exchange scenarios, a "strong logic-low cost" strategy is adopted: MT pre-translation utilizes "subject-customized NMT models" (e.g., "PubMed NMT model" for medical fields), automatically detecting "academic terminology consistency" and "logical connector usage" when processing abstracts. Error classification prioritizes identifying "logical coherence errors, academic terminology errors, and sentence structure confusion," while filtering out "colloquial expressions and non-critical data deviations." For proofreading prioritization: high-priority errors (e.g., research conclusion statements, experimental data translation deviations) are assigned to subject-matter translators; medium-priority errors (e.g., citation formatting) to general translators; low-priority errors (e.g., author affiliation translations) are automatically corrected by system templates. Cost control reduces proofreading effort for non-critical errors by 60% through "error filtering," keeping manual proofreading costs below $30 per thousand words.

Public service scenarios adopt a "high adaptability-broad coverage" strategy: MT pre-translation utilizes a "multi-language universal NMT model" to support translations for lesser-known languages (e.g., Vietnamese, Swahili). Upon text input, it loads a "target country cultural taboo database" (e.g., avoiding "pig"-related expressions in Saudi Arabia); Error classification prioritizes identifying "cultural conflict errors, colloquial adaptation errors, and missing critical information," while filtering out "minor semantic deviations and non- core vocabulary errors"; For proofreading prioritization: High-priority errors (e.g., mistranslations of "identity information" or "purpose of application" in visa documents) are assigned to bilingual cultural specialists; medium-priority errors (e.g., time expressions in public notices) are handled by regular translators; low- priority errors (e.g., politeness deviation) are automatically adjusted by the system based on the cultural database. For coverage assurance, in low-volume language scenarios, the "MT draft + light proofreading" approach (proofreading only high-priority errors) achieves a daily processing capacity of 100,000 words, meeting high-frequency public service demands.

IV. EXPERIMENTAL VALIDATION AND EFFECT ANALYSIS

A. EXPERIMENTAL DESIGN AND EVALUATION METRICS

Evaluation metrics are established across three dimensions: efficiency, quality, and user satisfaction.

Efficiency metrics include processing speed (number of words processed per unit time, words/hour), proofreading cost (average proofreading cost per thousand words, USD/k words), and response time (total time from text input to proofreading completion, hours). Quality metrics encompass accuracy rate (proportion of correctly translated key information, %, where key information is defined by domain experts), cultural adaptability (proportion of text free of cultural conflict errors, %, assessed by target-country cultural experts), and terminology consistency (proportion of consistent professional terminology translations within the same text, %). User satisfaction was measured via a 5-point scale questionnaire (1=Very Dissatisfied, 5=Very Satisfied), surveying users in three scenarios (business negotiators, researchers, public service applicants) regarding their satisfaction with translation results. This study was approved by the Ethics Committee of University of Sanya, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

Note: The 'Processing Speed' in this study refers to the collaborative processing speed of 'machine translation + human proofreading', not pure human proofreading speed. Machine translation undertakes over 90% of basic translation work, while humans only proofread 20%-40% of high-risk errors, resulting in significant overall speed improvement.

B. EXPERIMENTAL RESULTS AND VERIFICATION OF ADAPTABILITY TO LESS-COMMON LANGUAGES

Across all three scenarios, the optimized model group demonstrated significantly superior efficiency metrics compared to the traditional model group. In business negotiation scenarios, the optimized model achieved a processing speed of 4,500 words per hour—a 150% increase over the traditional model (1,800 words per hour). Proofreading costs dropped to $42 per thousand words (compared to $65 per thousand words for the traditional model), and response time

was reduced to 5.5 hours (vs. 22 hours for traditional mode). In academic exchange scenarios, the optimized mode processed at 5,800 words per hour—a 164% increase over the traditional mode (2,200 words per hour)—with proofreading costs at $26 per thousand words (traditional mode: $42 per thousand words) and response time at 22 hours (traditional mode: 68 hours). In public service scenarios, the optimized mode processes 8,200 words per hour—a 134% increase over the traditional mode (3,500 words/hour). Proofreading costs $13 per thousand words (traditional mode: $22/kW), with a response time of 6.8 hours (traditional mode: 20 hours). The efficiency gains primarily stem from:
- "Error filtering" reducing proofreading workload by 60%-70% - "Dynamic weight allocation" optimizing cost structure - "Parallel processing" shortening response time as detailed in Table 3 below.

TABLE 3. Comparison of Efficiency Indicators Between Two Modes

Scenario Type	Mode	Processing Speed (Words/Hour)	Proofreading Cost (USD/1,000 Words)	Response Time (Hours)
Business Negotiation	Traditional Mode	1800	65	22
Business Negotiation	Optimized Mode	4500	42	5.5
Business Negotiation	Improvement Rate	+150%	-35%	-75%
Academic Exchange	Traditional Mode	2200	42	68
Academic Exchange	Optimized Mode	5800	26	22
Academic Exchange	Improvement Rate	+164%	-38%	-68%
Public Service	Traditional Mode	3500	22	20
Public Service	Optimized Mode	8200	13	6.8
Public Service	Improvement Rate	+134%	-41%	-66%

Note: Improvement rate calculation logic: Improvement rate = (Optimized Mode Indicator - Traditional Mode Indicator) / Traditional Mode Indicator × 100%. For example, the processing speed improvement rate in business negotiation scenarios = (4500 - 1800) / 1800 × 100% = 150%, corresponding to 2.5 times the original speed.

To verify the model's generalization ability, a cross-domain test set was constructed, consisting of unseen cross-cultural texts covering legal, medical, and trade fields (not involved in any fine-tuning). The results show the model's error recognition F1-score on this test set reached 0.87, a decrease of only 0.05 compared to the training set, indicating no severe overfitting and good cross-scenario adaptability. After optimizing the professional term deviation recognition logic, the recognition accuracy increased from 82% to 93%, with the synonym misjudgment rate decreasing by 80% and the contextual mismatch omission rate decreasing by 75%.

The quality metrics of the optimized model group not only did not decline due to efficiency improvements but actually increased by "focusing on high-risk errors." In business negotiation scenarios, key information accuracy rose from 90% to 96%, cultural adaptability increased from 88% to 95%, and terminology consistency improved from 89% to 97%. In academic exchange scenarios, key information accuracy climbed from 85% to 92%, cultural adaptability advanced from 82% to 89%, and terminology consistency reached 94% from 86%. In public service scenarios, key information accuracy rose from 82% to 88%, cultural adaptability from 75% to 92%, and terminology consistency from 80% to 88%. These quality improvements stemmed from "intelligent error classification" ensuring 100% detection of critical errors, "priority proofreading for cultural conflicts" mitigating cultural misunderstandings, and "real-time terminology database matching" guaranteeing consistency, as detailed in Table 4 below.

TABLE 4. Comparison of Quality Indicators Between Two Modes

Scenario Type	Mode	Accuracy (Key Information)	Cultural Adaptability	Term Consistency
Business Negotiation	Traditional Mode	90	88	89
Business Negotiation	Optimized Mode	96	95	97
Business Negotiation	Improvement Rate	+6	+7%	+8%
Academic Exchange	Traditional Mode	85	82	86
Academic Exchange	Optimized Mode	92	89	94
Academic Exchange	Improvement Rate	+7	+7	+8%
Public Service	Traditional Mode	82	75	80
Public Service	Optimized Mode	88	92	88
Public Service	Improvement Rate	+6	+17	+8

User satisfaction surveys indicate significantly higher satisfaction in the optimized model group compared to the traditional model group: business scenario satisfaction rose from 3.2 to 4.6 points, academic scenario satisfaction increased from 3.0 to 4.5 points, and public service scenario satisfaction climbed from 2.8 to 4.7 points. Business users valued "terminological accuracy and legal unambiguity," academic users appreciated "logical coherence and scholarly standards," while public service users favored "cultural adaptation and information completeness." To validate the optimization framework's effectiveness in minor language scenarios, supplementary experiments were conducted on two minor language pairs: Arabic-Chinese (Middle Eastern business negotiations) and Swahili-Chinese (African public services). Results showed Arabic-Chinese business processing speeds reached 3,800 words/hour (vs. 1,500 words/hour for the traditional model), with proofreading costs of $48 per thousand words, 94% accuracy, and 96% cultural adaptation. For Swahili-Chinese public services, processing speed reached 7,200 words per hour, proofreading costs were $15 per thousand words, accuracy was 86%, and cultural adaptation was 93%. This demonstrates that the optimized model maintains high efficiency and quality in minor language

scenarios, as shown in Table 5 below.

TABLE 5. Efficiency and Quality Indicators in Minor Language Scenarios

Language Pair	Scenario Type	Processing Speed (Words/Hour)	Proofreading Cost (USD/1,000 Words)	Accuracy (%)	Cultural Adaptability (%)
Arabic-Chinese	Business Negotiation	3800	48	94	96
Swahili-Chinese	Public Service	7200	15	86	93

V. CONCLUSIONS

The core mechanism of collaborative efficiency optimization manifests in three aspects:

First, "error filtering" achieves a balance between efficiency and quality. By employing a "BERT error recognition model + scenario-based filtering rules," low-risk and non-core errors are filtered out. This allows human proofreaders to focus solely on addressing 20%-40% of critical errors, boosting efficiency by 2.3-2.6 times. Simultaneously, by concentrating resources on high-risk errors, quality improves by 6-17 percentage points. Second, "dynamic weighting" adapts to varying scenario demands. Dynamic weighting factors in the FAHP evaluation matrix ensure proofreading resources precisely match core scenario requirements. For instance, raising the "cultural adaptation sensitivity" weight to 0.4 in public service scenarios improves cultural adaptation by 17 percentage points. Third, "feedback iteration" creates a synergistic closed loop.

Short-term (monthly) HC correction data is used for MT model fine-tuning, reducing similar error rates by 30%-40%. Long-term (quarterly) updates to the error type library enhance model coverage, achieving a positive cycle: "MT quality improvement $\rightarrow$ reduced proofreading workload $\rightarrow$ further efficiency gains." The research has three limitations: First, insufficient coverage of less-commonly-taught languages—the current error recognition model only supports English, Japanese, Arabic, and Swahili, with accuracy below 75% for other languages like Urdu and Hausa. Second, limited adaptation to complex scenarios, excluding "multimodal cross-cultural communication" contexts (e.g., simultaneous interpretation assistance at international conferences, subtitle translation for short videos), which require integration of audio and visual information. Third, real-time performance needs improvement, with average processing time for error detection and priority ranking at approximately 2-3 minutes per thousand words, failing to meet the minute- level response demands of real-time scenarios (average response time for customer service scenarios is 3-5 minutes) (e.g., cross-border live-streaming customer service).

Future research can focus on four areas: First, expand multilingual models by incorporating "multilingual pre-trained models (e.g., mBERT, XLM-RoBERTa)" to enhance error recognition accuracy for lesser-used languages, aiming to cover 20+ languages commonly used in cross-cultural communication. Second, develop multimodal collaborative frameworks integrating speech recognition (ASR) and image recognition (OCR) technologies to extend the framework to scenarios like "speech translation + human proofreading," "image text translation + human proofreading," and optimize real-time performance through approaches such as "model lightweighting (e.g., distilling BERT models)" and "edge computing deployment." image-text translation + human proofreading" scenarios, such as real-time speech processing and proofreading push during international conferences; Third, optimize real-time performance by adopting "model lightweighting (e.g., distilling BERT models)" and "edge computing deployment" technologies to reduce error recognition time to under 10 seconds per thousand words; Fourth, industry-tailored solutions were designed for cross-border e- commerce, international healthcare, and diplomatic sectors. This involved developing "industry-specific error type libraries" and "proofreading weight models," such as prioritizing translation accuracy for "disease names and medication dosages" in international medical contexts, or similarly, for the critical technical parameters and material compositions found in manufacturing documentation.

AUTHOR CONTRIBUTIONS

Conceptualization –Zhaoxia Liu, Yingying Du and Xiaohui Zhang; methodology – Zhaoxia Liu, and Xiaohui Zhang; investigation – Zhaoxia Liu, Yingying Du and Xiaohui Zhang; writing-original draft preparation – Zhaoxia Liu, Yingying Du and Xiaohui Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research was supported by, 1. This paper is part of a series of research outputs from Hainan Provincial Philosophy and Social Sciences Planning Project (Foreign Language Base Special Project), project number: HNWYJD25-02, The research project titled "A Study on the Industry-Education Integration Talent Development Mechanism for 'Multilingualism + High-End Equipment Manufacturing."
2. This paper is also a part of research outputs from Research on the Optimization Path of Multilingual Government Service Platform in the Construction of Sanya Digital Government, Project Number: SYSK 2025-09.

HUMAN RESEARCH SUBJECTS

This study was approved by the Ethics Committee of University of Sanya, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] Y. Gao, F. Hou, and H. Jahnke, "Pre-ordering representations improve low-resource neural machine translation and application in the Māori language," Multimedia Tools and Applications, vol. 85, no. 1, 26, 2026, doi: 10.1007/S11042-026-21153-5.

[2] A. Tian and M. Sun, "Automatic Evaluation Method for Machine Translation Quality Based on Cross-Attention Recurrent Neural Network," Journal of Circuits, Systems and Computers, prepublish, 2026, doi: 10.1142/S0218126626500787.

[3] E. Satir and H. Bulut, "Controlled beam search for neural machine translation using subword units leveraging phrase-based statistical machine translation outputs," Discover Computing, vol. 29, no. 1, 37, 2026, doi: 10.1007/S10791-025-09633-Y.

[4] J. A. Vicente, F. T. Amamampang, D. D. Lahaylahay, and C. Cheng, "ChavacanoMT: a corpus and evaluation of neural machine translation for Philippine Creole Spanish," Language Resources and Evaluation, vol. 60, no. 1, 18, 2026, doi: 10.1007/S10579-025-09888-3.

[5] S. Song, Y. Chen, X. Hu, and J. Zhang, "Domain-Aware Transformer for Multi-Domain Neural Machine Translation," Computers, Materials & Continua, vol. 86, no. 3, 68, 2026, doi: 10.32604/CMC.2025.072392.

[6] Z. Man, Y. Zhang, Y. Chen, Y. Chen, and J. Xu, "DKF: Domain knowledge fusion in progressive incremental learning for multi-domain machine translation," Expert Systems With Applications, vol. 306, 130838, 2026, doi: 10.1016/J.ESWA.2025.130838.

[7] T. Alshaikhi, "AI-Integrated Translation Training and Translators' Competence System Using Neural Machine Translation (NMT)," SN Computer Science, vol. 7, no. 1, 60, 2026, doi: 10.1007/S42979-025-04661-3.

[8] R. Appicharla, K. S. Jha, A. Ekbal, and P. Bhattacharyya, "Maithilimt: Developing Multi-Domain Parallel Corpus for Hindi-Maithili Machine Translation," Language Resources and Evaluation, vol. 60, no. 1, 12, 2026, doi: 10.1007/S10579-025-09890-9.

[9] S. Zhang and C. Zhao, "Machine translationese of large language models: Dependency triplets, text classification, and SHAP analysis," PloS one, vol. 21, no. 1, e0339769, 2026, doi: 10.1371/JOURNAL.PONE.0339769.

[10] X. Lyu, J. Li, D. Wei, M. Zhang, S. Tao, H. Yang, et al., "Multiphase and Multitask Prompt Tuning for LLM-Based Context-Aware Machine Translation," IEEE transactions on neural networks and learning systems, 2025, doi: 10.1109/TNNLS.2025.3646848.

[11] A. Saulītis, "Evaluating multilingual digital resources: machine translation adoption and user satisfaction across six European countries," Language Resources and Evaluation, vol. 60, no. 1, 7, 2025, doi: 10.1007/S10579-025-09884-7.

[12] S. Wang, S. Shahir, and U. M. Ismail, "A painting art rendering system by deep learning framework and machine translation," Scientific reports, 2025, doi: 10.1038/S41598-025-34058-4.

[13] N. Ding and J. Cao, "A Comparative Dependency Analysis of Human Translation and Machine Translation: A Case Study of English translation of To Live," Studies in Linguistics and Literature, vol. 9, no. 4, 130, 2025, doi: 10.22158/SLL.V9N4P130.

[14] D. Munkova, L. Benkova, M. Munk, L. Benko, and P. Hajek, "Predictive modeling of error categories in English-Slovak machine translation using automatic evaluation metrics," Machine Learning with Applications, vol. 23, 100810, 2026, doi: 10.1016/J.MLWA.2025.100810.

[15] O. S. A. Elsayed, "When machines meet gavel: a case study of the English–Arabic machine translation of the Egyptian arguments before the International Court of Justice (2024)," Language and Semiotic Studies, vol. 11, no. 4, pp. 661-690, 2025, doi: 10.1515/LASS-2025-0054.

[16] Z. Yang, "Investigating Undergraduate Students' Perceptions of Post-editing Efforts in Translation Memory and Neural Machine Translation," Journal of Humanities, Arts and Social Science, vol. 9, no. 11, pp. 2200-2206, 2025, doi: 10.26855/JHASS.2025.11.022.

[17] Q. Zhao, J. Guo, and Z. Yu, "Recursive mutual-supervised multimodal mask-variational fusion with domain-anchor alignment for domain-specific Multimodal Neural Machine Translation," Engineering Applications of Artificial Intelligence, vol. 165, 113365, 2026, doi: 10.1016/J.ENGAPPAI.2025.113365.

[18] B. Huang, M. Mao, and Y. Wang, "Non-autoregressive Machine Translation with Target Language Components in English Corpus," Journal of Information & Knowledge Management, 2550120, 2025, doi: 10.1142/S0219649225501205.

[19] A. Sindhujan, D. Kanojia, and C. Orăsan, "Reference-Less Evaluation of Machine Translation: Navigating Through the Resource-Scarce Scenarios," Information, vol. 16, no. 10, 916, 2025, doi: 10.3390/INFO16100916.

[20] K. Welnitzová and D. Munková, "Persisting Grammatical Errors of Machine Translation," Journal of Linguistics/Jazykovedný casopis, vol. 76, no. 2, pp. 468-492, 2025, doi: 10.2478/JAZCAS-2025-0039.

[21] S. Gou, "Research on the Optimization Path of Machine Translation from a Cognitive Perspective: The Collaborative Development of Human Translation and Machine Translation," The Frontiers of Society, Science and Technology, vol. 7, no. 7, 2025, doi: 10.25236/FSST.2025.070709.

[22] S. Sheikh, Y. Dongkeun, C. Jiwoo, J. Woori, and J. Seohyon, "Evaluating English-Korean Literary Machine Translations: A Dataset Featuring the RULER and VERSE Annotation Methods," Journal of Open Humanities Data, vol. 11, no. 1, 62, 2025, doi: 10.5334/JOHD.393.

[23] J. Huarui and H. Shuai, "Neural Machine Translation for Multilingual Human–Machine Collaboration in Smart Factories Supporting the IIoT," Internet Technology Letters, vol. 9, no. 1, e70178, 2025, doi: 10.1002/ITL2.70178.

[24] Y. Li, J. Ye, and J. Guo, "Two-stage incremental semantic aggregation for domain-specific multimodal neural machine translation," Expert Systems With Applications, vol. 299, 130086, 2026, doi: 10.1016/J.ESWA.2025.130086.

[25] N. L. Santiáñez and C. G. Pastor, "Measuring Creative Phraseology in Literature: Machine Translation Systems Versus Large Language Models," Yearbook of Phraseology, vol. 16, no. 1, pp. 125-152, 2025, doi: 10.1515/PHRAS-2025-0006.

[26] V. T. N. Phuong, T. P. Huong, and P. P. Lan, "Integrating AI-Based Machine Translation in Translation Pedagogy: Evidence from a Mixed-Methods Study in Vietnam," Education, Language and Sociology Research, vol. 6, no. 4, 37, 2025, doi: 10.22158/ELSR.V6N4P37.

[27] B. Wang, X. Hui, and S. Jin, "Machine Translation Constraints in Lexical Conversion in Classical Chinese and Implications for China's English Teaching," Journal of Education and Educational Research, vol. 15, no. 3, pp. 27-31, 2025, doi: 10.54097/2PAST304.

[28] M. Sasaki, A. Mizumoto, and K. P. Matsuda, "Machine translation as a form of feedback on L2 writing," International Review of Applied Linguistics in Language Teaching, vol. 63, no. 4, pp. 2301-2326, 2025, doi: 10.1515/IRAL-2023-0223.

[29] X. Li, X. Wang, and W. Lai, "The Usability of Neural Machine Translation in Creative-Text Post-Editing: Evidence from Users' Performance and Perception," International Journal of Human–Computer Interaction, vol. 41, no. 21, pp. 13792-13803, 2025, doi: 10.1080/10447318.2025.2476714.

[30] M. R. Alhubayshi, M. S. Fallata, and A. N. Alowedi, "Examining the Quality of Machine Translation Subtitling for Saudi Series: Tash Ma Tash a Model," Forum for Linguistic Studies, vol. 7, no. 10, pp. 36-46, 2025, doi: 10.30564/FLS.V7I10.10617.

[31] H. Jin, "A Comparative Analysis of the Cognitive Processes in Machine Translation and Human Translation," Journal of Modern Educational Theory and Practice, vol. 2, no. 6, 2025, doi: 10.70767/JMETP.V2I6.717.

[32] Y. Li and X. Song, "Beyond Neutrality: Mapping Two Decades of Research on Machine Translation Bias (2005–2024)," SAGE Open, vol. 15, no. 4, 2025, doi: 10.1177/21582440251392700.

[33] C. Prichard and A. Atkins, "Variables affecting the use and outcomes of machine translation for reading: Text difficulty and cognitive reading anxiety," System, vol. 134, 103822, 2025, doi: 10.1016/J.SYSTEM.2025.103822.

[34] P. J. Jayan, S. J. Kumar, and T. Amudha, "Challenges and improvisation in machine translation: the case of malayalam–tamil machine translation," Language Resources and Evaluation, vol. 59, no. 4, pp. 3478-3520, 2025, doi: 10.1007/S10579-025-09840-5.

[35] M. Gupta, M. Dutta, and K. C. Maurya, "Direct speech-to-speech neural machine translation: A survey," Speech Communication, vol. 175, 103317, 2025, doi: 10.1016/J.SPECOM.2025.103317.