European Journal of Multilingualism and Translation Studies
ISSN: 2734 - 5475
ISSN-L: 2734 - 5475
Available on-line at: www.oapub.org/lit

DOI: 10.46827/ejmts.v6i1.718 Volume 6 | Issue 1 | 2026

AI VERSUS HUMAN ENGLISH-VIETNAMESE TRANSLATION: A STUDY AT A PRIVATE UNIVERSITY

Bùi Kim Phượng[i],
Nguyễn Minh Trang
Bình Dương University,
Ho Chi Minh City,
Vietnam

Abstract:
The rapid progress in Artificial Intelligence (AI) technology, particularly in Neural Machine Translation (NMT) systems, has significantly impacted translation practices across the world. The use of AI systems like ChatGPT and DeepL is becoming more popular, and young translators in translation programs in Vietnam are more likely to use these systems in their academic work. However, there is little research on the pedagogical implications of these systems. This research focused at examining the effectiveness of AI-assisted translation and human translation in English-Vietnamese language pair for 67 English major students at Binh Duong University. A mixed-methods design was employed, in which two translation tasks were assigned: independent human translation and AI-assisted translation followed by post-editing to discover the two research questions. An analytic rubric was used in assessing translation quality, and descriptive statistics and thematic analysis were employed in data analysis. The results show that AI-assisted translation is superficial in nature, as only semantic accuracy is fulfilled while terminology, cultural sensitivity and pragmatic ability are lacking. The AI tool scored on average Fair to Good, while human translation scored Very Good on all nine criteria. Novice translators over-relied on the AI tool and lacked the critical thinking required in the post-editing phase, although they recognized its importance. More than half also expressed the need for pedagogical support. The study indicates that the AI tool should be used as a supplement, not a substitute, in the pedagogical process and recommends that teaching AI literacy, critical review, fast engineering, and post-editing should be done along with human translation.

Keywords: AI-assisted translation, human translation, neural machine translation, post-editing, novice translators

[i] Correspondence: email phuongbinhduongbui@gmail.com

1. Introduction

1.1 Background of the study

The digital revolution has significantly impacted many professional fields, and translation is one such field that has undergone a tremendous change. Since the breakthrough attention mechanism proposed by Bahdanau *et al.* (2015), Neural Machine Translation systems have supplanted previous rule-based and statistical methodologies owing to their enhanced capacity for contextual processing. The following advancement of the Transformer architecture (Vaswani *et al.*, 2017) and Large Language Models (LLMs) like ChatGPT significantly enhanced AI translation capabilities to unparalleled heights. The emergence of AI-assisted translation tools, including massive language models and neural machine translation systems like DeepL, Google Translate, and ChatGPT, has profoundly influenced translation methods. In Vietnam, currently AI-powered translation apps are commonly used in universities. Students often make use of ChatGPT and DeepL to fulfill their assignments, especially translation assignments. This is where the quality of AI translation emerges as an issue, especially in terms of translation competence.

1.2 Rationale of the Study

English and Vietnamese languages vary significantly in terms of grammar, pragmatics, and sociolinguistics. For instance, the grammar of the English language is normally Subject-Verb-Object, and it is an inflected language. On the other hand, the Vietnamese language is analytic, and its meaning depends greatly on tone. Thus, these significant differences may cause the two languages operating differently, particularly in translation practices.

1.3 Research Objectives

The research seeks to evaluate the translation quality of AI versus human translators for accuracy, fluency, and cultural relevance, investigate the views of English students on AI translation products, including advantages and disadvantages, and AI potential use in language education and propose translation strategies for the application of AI translation at higher educational institutions.

1.4 Research Gap

Scholars have examined the discourse around machine translation in contrast to human translation (Arnold *et al.*, 1994; Koehn, 2020; O'Brien, 2012; Pym, 2013). Recent research includes an examination of the translation quality of ChatGPT (Jiao *et al.*, 2023). In the Vietnamese context, Nguyen *et al.* (2022) proposed the Vinai translate model, which achieved the state-of-the-art performance in Vietnamese-to-English neural machine translation. nevertheless, the majority of the research is centered on the performance of the translation tools for the work of professional translators.

In terms of the application of the translation tools among undergraduate English majors, the empirical data is still scarce globally and totally absented in the Vietnamese context (Jiao *et al.*, 2023; Nguyen *et al.*, 2022). moreover, the language pair under investigation is the

Vietnamese English pair, which is considered to be a distant language pair with significant syntactic, pragmatic, and cultural differences (Pym, 2013; Koehn, 2020).

This study is designed to address the aforementioned gaps in the existing research by focusing on the application of the translation tools among the third-year English majors at Binh Duong University and the framework for the application of the translation tools in translation teaching

1.5 Research Questions

This study is motivated by two research questions:

RQ1: How does AI translation compare with Human Translation in terms of accuracy, fluency, and cultural appropriateness among English majors at Binh Duong University?

RQ2: What are English majors students' perceptions of the application of AI translation in their learning process?

The contextual framework and research questions outlined in this section will form the basis for the analytical framework for the following literature review. Section II will explore the theoretical and empirical basis for the study, in relation to the existing theories of translation, the latest advances in Neural Machine Translation, and novice translator cognition.

2. Literature Review

2.1 Human Translation

Human Translation (HT) is a highly complex cognitive and linguistic process that goes far beyond the simple substitution of one word for another. Bell (1991) and House (2015) have described HT as a highly complex process requiring expert handling of linguistic, pragmatic, and cultural variables. This expertise is often rooted in what is termed *'linguistic intuition*'. A cognitive skill that allows human translators to detect subtle subtleties and natural language that today's AI machines sometimes fail to grasp. Jakobson's (1959; 2004) theory of intersemiotic translation implies that human translators do not translate individual words, but interpret signs in a complex way, keeping meaning at several levels of grammar and cultural complexity.

In the work of Mammadova (2024) it has been emphasised the role of HT as a rigorous science of structure and an art of careful rebuilding. The study describes a difficult four-stage process of HT: text decoding, cultural features analysis, post-editing transfer and review, each of which requires a certain level of discernment that current AI models are unable to perform. The most unique ability of human translators lies in their capacity to handle domain-specific structures, context-dependent cultural values, and terminology in critical domains like medicine, law, and engineering, as highlighted in the works of House (2015).

2.2 Machine Translation and AI-Assisted Translation

2.2.1 Neural Machine Translation (NMT)

The history of machine translation can be viewed as an evolutionary process from the initial dreams of achieving fully automatic high-quality translation (FAHQT) in the 1950s to the more

pragmatic approach of human-machine collaboration. The major innovation in machine translation is the invention of the neural machine translation approach. According to Koehn (2020), the sequence-to-sequence architecture with the use of deep neural networks is the major innovation in machine translation, where the networks learn from the data rather than relying on rules. The application of the attention mechanism by Bahdanau *et al.* (2015).

Bahdanau *et al.* (2015) proposed the attention mechanism, which solved the problem of phrase length constraints and increased the equivalence of source and target sentences. The self-attention mechanism, introduced by Vaswani *et al.* (2017) with the invention of the transformer, revolutionised the approach of neural machine translation by enabling parallel processing and information retrieval from Modern LLMs such as ChatGPT that have adopted this approach and improved it by increasing the context window to process multiple texts for translation. However, the output of machine generation is not mistake free and cannot be entirely relied on (Koponen, 2016), hence post processing is a crucial activity for quality control. as noted by Zouhar *et al.* (2021). Bahdanau *et al.* (2015) proposed the attention mechanism, which solved the problem of phrase length constraints and increased the equivalence of source and target sentences. The self-attention mechanism, introduced by Vaswani *et al.* (2017) with the invention of the transformer, revolutionised the approach of neural machine translation by enabling parallel processing and information retrieval from Modern LLMs such as ChatGPT that have adopted this approach and improved it by increasing the context window to process multiple texts for translation. However, the output of machine generation is not mistake free and cannot be entirely relied on (Koponen, 2016), hence post processing is a crucial activity for quality control. as noted by Zouhar *et al.* (2021), the effectiveness of post-editing is significantly dependent on the initial quality provided by the NMT systems.

2.2.2 AI-Assisted Translation and Its Limitations

In the current Large Language Model (LLM) landscape, the role of AI in the translation process is not only more refined but also involves an enhanced understanding of the context. The architectural improvement in the performance of AI-based machine translation tools is accompanied by the occurrence of certain unique types of errors. In a comparison study of ChatGPT and Google Translate, Al-Afnan (2024) states that LLMs possess flexible language abilities, but they have unique issues regarding computational consistency across specific tasks. Furthermore, according to Wei *et al.* (2025), the phenomenon of "representation drift" is observed, where the "cumulative noise" in the context window results in the AI system producing "self-consistent hallucinations" that are not only semantically different from the source text but also "internally fluent" and "logical." This phenomenon highlights the fact that the "self-attention" mechanism introduced by Vaswani *et al.* (2017) is not effective in the prevention of the fabrication of content.

In addition to the above technical limitations, the role of the psychological aspect is also recognized by Martindale and Carpuat (2018) as fluency is found to play a greater role than accuracy in the judgment made by the user. This phenomenon is referred to as the "trust trap," where the novice translator is trapped while the AI system focuses on the "Zero-shot learning" mechanism, as discussed by Wolff (2024).

2.3 Post-Editing and Novice Translators

Machine Translation Post-Editing (MTPE) is described by Williams (2013) as a process by which translators refine raw machine translation output to reach publication quality standards. The cost-effectiveness of machine translation post-editing is seen to be contingent on the quality of machine translation output. When post-editing is seen to take longer than translating from scratch, the economic viability of machine translation post-editing is negated (Koehn, 2020).

MTPE is also seen to pose significant cognitive challenges for the novice translator. Wei *et al.* (2025) and Martindale and Carpuat (2018) highlight two major challenges for the novice translator. The first is the 'self-consistent hallucination' trap, wherein the machine translation output is logically consistent but incorrect. The second is the 'attention-locking' phenomenon, wherein the machine translation output is locked into an incorrect pathway and is cognitively costly for the post-editor. The challenges posed by machine translation post-editing highlight the need for the novice translator to not only possess linguistic competence but also the competence to critically assess machine translation output – a competence described by Hutchins (1995) and Williams (2013) as 'human-machine symbiosis competence.'

2.4 Translation Theories

2.4.1 Equivalence Theory

Eugene Nida's Equivalence Theory (Nida, 1964) argues that formal equivalence, which prioritizes the formal correspondence between source and target texts, and dynamic equivalence, which prioritizes the effect of the translation on the target audience, are two different but distinct translation strategies. This theory is the basis on which the accuracy of the translation will be measured in this study: Even if formal correspondence is achieved in the translation carried out by the AI, it may not result in dynamic equivalence in the cultural context.

2.4.2 Skopos Theory

Skopos Theory, developed by Hans Vermeer (1989), suggests that the *skopos* (purpose) of the target text dictates the mode of translation. Thus, to assess the quality of the translation, it is necessary to determine the intended *skopos* of the target text. For the academic texts intended to be translated into Vietnamese and read by university students in Vietnam, the *skopos* will be to provide an accurate and culturally appropriate translation, ensuring that the communicative function is maintained for the target audience.

2.4.3 Text Typology

According to (Reiss, 1971), Katharina Reiss's theory of text typology, in which texts are classified into informative, expressive, and operative texts, will also influence the selection of translation strategies. The academic source text will be classified as informative.

2.5 Translation Strategies

Baker and Saldanha (2020), citing Lörscher (1991), characterize translation strategy as a "*potentially conscious process aimed at solving a translation problem*." The PACTE group (2011)

also highlights the role of strategy in translation competence and characterizes it as a process comprising four consecutive phases: problem identification, resource mobilization, strategy selection and application, and monitoring and evaluation. This process is particularly relevant to the use of AI tools in translation, as a translator who is not aware of a translation strategy will essentially bypass the problem-identification and monitoring phases of the process and will not monitor the quality of the final product for pragmatic and cultural appropriacy.

3. Methodology

3.1 Research Design

This research employed a convergent parallel mixed-methods design to address the two research questions posed in this study, as proposed by Creswell and Creswell (2018). This design employed quantitative and qualitative data to address the research questions. Quantitative data were collected in the form of translation quality data and Likert scale data, while qualitative data were collected in the form of semi-structured interview data.

This research design was conducted in two parallel phases to address the research questions. Phase 1 of this research design was employed to address Research Question 1 and focused on collecting and analyzing translation quality data. Phase 2, which served purpose for RQ2, entailed the collection of perception data using the survey questionnaire and subsequent interviews with the subsample participants.

3.2 Participants and Sample

The participants were English major students from the Faculty of Foreign Languages, Binh Duong University. The reason behind the selection criteria is that the students had completed at least one English-Vietnamese translation course and had experience with at least one AI translation tool, such as Google Translate, DeepL, ChatGPT, etc.

The sample size consists of 67 English major students from the Faculty of Foreign Languages, Binh Duong University, for the purpose of conducting the translation task and the questionnaire survey. The sample size is adequate as it meets the requirements of the following methods: Cohen's (1988) requirement for the sample size, which is 50-64 for the purpose of detecting medium effects with t-tests with 0.80 power; Field's (2013) requirement for the sample size to be greater than 50 for the purpose of ensuring means in hypothesis testing; Tabachnick and Fidell's (2013) requirement for the sample size to be greater than or equal to 50 for the purpose of ensuring Cronbach's Alpha in scale reliability assessment. Eight students were purposively selected from 67 students for interviews, as Guest *et al.* (2006) reported data saturation in qualitative translation perception research is achieved in 6-12 interviews. The criteria used in selecting interviewees were based on their usage of AI as rare, occasional, and frequent users.

Table 3.1: Overview of Research Design and Participant Groups

Participant Group	**Method**	***N***	**Purpose**
Primary sample	Translation task & Survey questionnaire	67	Quantitative (RQ1 & RQ2)
Subset-interview group	Semi-structured interviews	8	Qualitative (RQ2)
Expert English-Vietnamese raters *External validators*	Blind evaluation of translations	2	Validation Translation quality

Note that the quantitative strand of the study (N=67) targeted the research questions using task-based data and self-report measures. The qualitative subset (n=8) was purposively selected from the primary sample to provide further information related to the results of RQ2. The expert raters worked independently to guarantee the validity of translation quality.

Table 3.2: Demographic Profiles and AI Usage Frequency of Participants (N=67)

Characteristics	**Categories**	**Frequency (*f*)**	**Percentage (%)**	**Mean (*M*)**	**Std. Deviation (*SD*)**
Gender	Male	21	31.3	1.69	0.47
	Female	46	68.7		
Major	English studies	67	100.0	1.00	0.00
Year of Study	First year	19	28.4	2.45	0.94
	Second year	1	1.5		
	Third year	45	67.2		
	Fourth year	2	3.0		
AI Usage Frequency	Sometimes	64	95.5	3.04	0.21
	Often	3	4.5		

As presented in Table 3.2, the research sample consisted of 67 participants who were characterized by absolute homogeneity in their academic major (M=1.00, SD=0.00). Female students were the majority in the research sample (68.7%, n = 46). Third-year students were the primary group in the research sample (67.2%, n=45). The mean frequency score of AI usage was 3.04 (SD=0.21), indicating that 95.5% of the participants used AI tools only "sometimes" in their academic and translation work.

3.3 Data Collection Instruments

3.3.1 Translation Task

A 192-word English-language academic source text was adopted as a tool for translation (*see Appendix 2*). This text was specifically crafted to incorporate three types of translation difficulties, each of which corresponds to one of the three evaluation criteria. For evaluation on accuracy, specific vocabulary such as "pragmatic competence," "discourse analysis," and "contextual inference" was included, as these terms require more than literal translation. For evaluation on fluency, complex sentence structures such as 25- to 30-word sentences, relative clauses, parallel structures, and subordinate clauses were also included. An idiomatic expression such as "a blessing in disguise", distinctions in English and Vietnamese discourse styles and a cultural concept such as "relational harmony" were also added for evaluation on

cultural appropriateness. This text has been pre-checked by translation professionals for its suitability and difficulty level.

The translation task was conducted in two phases: Phase 1, where no AI tool was used, took 20-25 minutes, and Phase 2, where an AI tool was used, took 15-20 minutes, with a one-week gap to avoid practice effects.

3.3.2 Analytical Scoring Rubric

All the translations were assessed using an "Analytical Scoring Rubric" based on various quality assessment models for translation (*see Appendix 4*), such as those developed by O'Brien (2012), House (2015), and Nababan *et al.* (2012). The rubric was graded on a scale of 1 to 5, with 1 being "Poor", 2 being "Weak", 3 being "Fair", 4 being "Good" and 5 being "Excellent". The parameters taken into consideration were: "Accuracy", "Fluency" and "Cultural Appropriateness". 'Accuracy' meant 'faithful delivery of information contained in the original message without adding, deleting or distorting information, especially with regard to terms peculiar to a certain field or profession'. "Fluency" referred to "grammatical correctness, naturalness, and coherence of the target text written in Vietnamese." "Cultural Appropriateness" referred to "successful adaptation of idiomatic expressions, register, and cultural references to an academic context in Vietnam." Blind evaluation was conducted, where raters were not informed that they were to evaluate AI-generated or human-produced texts.

3.3.3 Survey Questionnaire

A 32-question survey questionnaire was designed to collect data about students' perceptions of AI in translation learning (*see Appendix 1*). The structure of the questionnaire was divided into two sections. Section 1 contained 6 questions to collect students' demographic and background information, including their usage of AI tools, frequency of usage, and purpose of usage. Section 2 consisted of 26 questions, including 24 five-point Likert scale questions ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), which were grouped into four major themes: (a) students' perceptions of the quality of AI translation compared with human translation; (b) students' perceptions of the benefits of AI in translation learning; (c) students' perceptions of dependency and reliance on AI; and (d) students' perceptions of pedagogical expectations and integration of AI. Two open-ended questions were designed to ask students to describe what they thought were the most important strengths and limitations of AI in translation learning. The questionnaire was piloted among five students who were not in the main sample, and some modifications were made to the questionnaire based on the feedback from the pilot students.

3.3.4 Semi-Structured Interview Protocol

A semi-structured interview guide consisting of 13 questions was designed, in line with Dörnyei (2007) and Kvale and Brinkmann (2015)'s recommendations on qualitative research on learner cognition (*see Appendix 3*). The three areas of thematic investigation were: benefits of using AI in translation learning, constraints and challenges faced, and suggestions on how to effectively incorporate AI in translation courses. Each interview was conducted in

Vietnamese, lasting 15-20 minutes, with full transcription. Eight participants were interviewed, in line with Guest *et al.*'s (2006) data saturation principle, and are referred to with pseudonyms P1, P2, P3, etc., in this analysis.

Table 3.3: Overview of Data Collection Instruments

Instrument	Content	Research Question
Source text (English)	192 words; specialized terms, complex syntax, idioms	RQ1
Scoring Rubric	5-point Likert; 3 criteria: Accuracy, Fluency, Cultural Appropriateness	RQ1
Survey Questionnaire	32 items (6 demographic, 24 Likert, 2 open-ended)	RQ2
Interview Protocol	13 semi-structured questions on AI strengths, limitations, pedagogy	RQ2

Note that all measures were piloted with a small sample of participants prior to administration. The scoring rubric was expert-validated prior to administration.

3.4 Data Collection Procedure

The data collection process involved a six-step procedure. Step 1 included selecting and validating the source text by experts. Step 2 entailed collecting Phase 1 translations under controlled conditions by humans only. Step 3 involved collecting Phase 2 translations one week later. Step 4 included arranging a blind evaluation by two experts in Phase 2 translations. Step 5 involved administering a survey questionnaire immediately after both translation phases. Step 6 included conducting semi-structured interviews with eight purposively selected participants.

3.5 Data Analysis

3.5.1 Quantitative Analysis

Quantitative data processing was employed using SPSS version 26. Two types of analysis were carried out. Descriptive statistics, including means, standard deviations, and frequency distributions, were computed for all the items included in the questionnaire to determine the perceptual trends. A paired t-test was applied to compare the means for the products developed using the AI-assisted translation tool and the human translation tool with respect to each evaluation criterion at $\alpha = 0.05$ significance level. Cronbach's Alpha (Nunnally, 1978) was employed to determine the scale reliability, while the inter-rater reliability was ensured using Cohen's Kappa (Landis & Koch, 1977), with the threshold set at $\kappa \geq 0.70$.

3.5.2 Qualitative Analysis

The data from the interviews have been analyzed through Thematic Analysis following the six-phase framework of Braun and Clarke (2019). The six-phase framework of Braun and Clarke (2019) for Thematic Analysis consists of familiarization, initial coding, theme searching, theme reviewing, theme defining, and report production. A directed coding

framework has been developed a priori (T1-T4) on the basis of the questions asked in the interview.

Table 3.4: Qualitative Coding Framework

Code	Theme	Content	RQ Link
T1	Accuracy & Terminology	Faithfulness to source; handling of specialized terms; AI hallucination	RQ1 & RQ2
T2	Fluency & Style	Naturalness of target language; syntax; cohesion; mechanical output	RQ1 & RQ2
T3	Cultural Appropriateness	Idiom handling; localization; Vietnamese socio-cultural norms	RQ1 & RQ2
T4	Benefits, Limitations & Pedagogy	Student perspectives on AI role, dependency, and instructional needs	RQ2

The codes were inductively generated from interview data and deductively generated from the theoretical framework. Approximately 20% of the data underwent independent coding by a second researcher, and Cohen's Kappa of 0.676 ($p < 0.001$) revealed substantial agreement between coders (Landis & Koch, 1977).

3.5.3 Data Integration and Triangulation

Data integration employed triangulation of the independent quantitative and qualitative data analysis results, as described by Creswell and Creswell (2018). Qualitative data provided contextualization and interpretation of quantitative data, particularly in instances of statistically significant differences in performance and polarized survey data distributions.

3.6 Research Ethics

The research was conducted in accordance with existing guidelines on ethics in educational research. All participants were required to give their informed consent, which included their consent to record their voices during the interview process. The research was conducted on a voluntary basis, and participants were free to withdraw their participation at any stage without any academic implications. All personal identifiable information was removed from the data processing stage, and participants in the interview were identified using pseudonyms, e.g., P1, P2, etc.

The research was conducted in accordance with UNESCO's Recommendation on AI Ethics (2021) and OECD's AI in Education guidelines (2022), where there was no usage of participant data for any training or testing of AI models in this research.

4. Results and Findings

This chapter presents the findings of the study based on the two research questions that were stated in the introduction. Data were obtained via three instruments: a 23-item survey questionnaire (N=67), semi-structured interviews (N=8), and a two-part translation assignment utilising an Analytical Scoring Rubric. Descriptive statistics and paired sample t-

test were used to examine quantitative data, while theme analysis was performed on qualitative data. You have the complete story on findings from two strands of data.

4.1 Preliminary Analysis: Reliability and Demographics

4.1.1 Scale Reliability

The internal reliability of the 24-item perception questionnaire was determined using Cronbach's Alpha (Appendix 5). Preliminary analysis revealed an α of 0.882 (standardized α of 0.887). This was an indication of a high reliability coefficient (Nunnally, 1978). An item-total correlation (CITC) analysis was performed, and we discovered that the item "AI fails to convey cultural nuances" had a critically low CITC of 0.169, much lower than the tolerable 0.30. Internal reliability was improved by dropping this item from the questionnaire. The α of the 23-item questionnaire was 0.884 after elimination, and the CITC of all remaining items was >0.30, ranging from 0.296 to 0.693. The item "AI correctly translates terminology" was retained with a borderline CITC of 0.296 because of its apparent relevance as a construct.

The 18-item paired comparison scale for translation quality revealed an excellent internal reliability of α = 0.984 (standardized α of 0.985), with all items having a high item-total correlation ranging from 0.835 to 0.927. Alpha-if-Item-Deleted was consistent, ranging from 0.983 to 0.984, indicating that all items contributed equally to the reliability of the scale. It should be noted that an alpha of 0.984 may indicate that there is the potential for item redundancy between the AI-Human items, which assess the same criterion, and is expected in the structure of the paired scales within the criterion, and does not affect validity. The inter-rater reliability of the translation scoring system was established with Cohen's Kappa = 0.676 ($p < .001$), indicating substantial agreement (Landis & Koch, 1977).

Table 4.1: Scale Reliability Summary

Scale	Items	Cronbach's Alpha	Assessment
Perception Questionnaire (initial)	24	0.882	Good (Nunnally, 1978)
Perception Questionnaire (after item removal)	23	0.884	Good
Translation Quality Comparison Scale	18	0.984	Excellent
Inter-rater Reliability (Cohen's Kappa)	—	0.676 (p< .001)	Substantial agreement (Landis & Koch, 1977)

4.1.2 Item-Level Reliability Patterns

Out of the retained perception items, three items exhibited exceptionally high CITC values, which further validated their importance in the construct. These items include "AI is a valuable tool" (CITC=0.690), "Using AI improves my translation competence" (CITC = 0.661), and "AI improves my productivity" (CITC=0.598). The items related to AI's management of ambiguous meanings (CITC=0.332), contextual incorrectness (CITC=0.310), and idiomatic expressions (CITC=0.359), although relatively lower, still indicate acceptable values and can be explained by the complexity of the construct, which encompasses quality skepticism, benefit perception, and dependency, rather than the reliability of the measurement tool.

For the 18-item translation quality scale, when the CITC values of the AI and human translation items are compared, it is noticeable that there is a consistent pattern of values in which the human translation items exhibit higher CITC values than their corresponding AI translation items in most criteria. The highest differences in CITC values are noticed in Cohesion & Coherence (Human=0.927, AI=~0.89), Idioms & Metaphors (Human= ~0.92, AI = ~0.88), and Cultural Sensitivity (Human= ~0.91, AI= ~0.87).

Table 4.2: CITC Comparison: AI vs. Human Translation Items

Criterion	CITC (AI)	CITC (Human)	Direction
Cohesion & Coherence	~0.89	0.927	Human higher
Idioms & Metaphors	~0.88	~0.92	Human higher
Cultural Sensitivity	~0.87	~0.91	Human higher
Register & Formality	~0.86	~0.88	Human slightly higher
Semantic Accuracy	~0.86	~0.87	Near equivalent
Grammatical Correctness	0.835	~0.84	Near equivalent

4.1.3 Descriptive Statistics: Student Perception Questionnaire

The descriptive statistics of the 23-item questionnaire showed that the mean scores ranged from 2.80 to 3.79 among the total of 67 participants, with no missing data points (*see Appendix 10*). The student perceptions were generally positive but somewhat neutral about the AI tools, realizing their applicability but showing some reservations about the linguistic depth. Four clusters were identified based on the pattern of the mean scores.

Table 4.3: Descriptive Statistics: Mean Scores by Thematic Cluster

Item Group	Mean Range	Dominant Trend
AI translation quality (terminology, idioms, cultural nuance)	2.81–2.94	Low skeptical of AI linguistic depth
AI benefits (efficiency, productivity, valuable tool)	3.46–3.75	High positive attitude toward AI utility
AI dependency (rely on, prefer, struggle without AI)	2.93–3.36	Moderate task-contingent reliance
Pedagogy & integration (guidance, curriculum needs)	3.33–3.51	Positive desire for structured integration

The quality skepticism cluster (M = 2.81–2.94) shows that participants do not believe that AI is effective in dealing with the complexities of the linguistic requirements of English-Vietnamese translation. The benefits cluster (M=3.46–3.75) shows that participants believe in the value of AI, which is a contradictory perception that is in line with all participants' consistent statements on AI being valuable but limited. The dependency cluster (M=2.93–3.36) shows that participants moderately rely on AI, but not entirely. The pedagogy cluster (M= 3.33–3.51) shows that students are ready for AI if appropriate pedagogical tools are provided.

4.1.4 Notable Frequency Results

The frequency analysis of each question produced six significant findings that will be used as a foundation for further qualitative exploration, which represent the strongest points of

agreement, positive or negative, on participants' perceptions of AI in translation. (*see Appendix 8,9*).

Table 4.4: Notable Frequency Results from the Perception Questionnaire (N = 67)

Survey Item	**Key Frequency Finding**	**Interpretation**
AI produces contextually incorrect translations	Agree + Strongly Agree: 70.1%	Most widely endorsed AI weakness
AI translation sounds mechanical	Agree + Strongly Agree: 68.6%	Strong consensus on fluency limitation
AI is a valuable tool (M = 3.75)	Agree + Strongly Agree: 67.1%	Highest positive rating in the entire dataset
Human translators adapt better to Vietnamese culture	Agree + Strongly Agree: 64.2%	Human cultural advantage is clearly perceived
I need guidance on AI use in translation (M = 3.51)	Agree + Strongly Agree: 50.8%	Clear and actionable pedagogical demand
AI correctly translates specialized terminology	Disagree: 26.9%	Notable skepticism about AI precision

The co-occurrence of the highest positive rating ('AI is a valuable tool,' 67.1% agreement) and the two highest negative ratings ('contextually incorrect,' 70.1%; 'sounds mechanical,' 68.6%) summarizes the paradox of the central theme of student AI perception in this study: AI is seen as being both useful and useless. The combination of both perceptions serves as the basis for the pedagogical recommendations proposed in Chapter 5. The disagreement rate of 26.9% for the accuracy of AI's terminology, the highest disagreement rate for any question, reinforces the skepticism of the student sample toward precision-based translation, as expected from the CITC pattern.

4.2 Research question 1: AI-assisted vs. human translation quality

To elicit answers for RQ1, a two-part translation task was carried out. The quality of both products was rated by two raters, each applying the Analytical Scoring Rubric for the three criteria of accuracy, fluency, and cultural appropriateness.

4.2.1 Accuracy

Accuracy was evaluated on three sub-criteria: semantic accuracy, terminology, and preservation of technical detail. Table 4.5 displays the mean values for both translation conditions.

Table 4.5*:* Mean Scores: Accuracy Criteria (AI vs. Human Translation)

Criterion	**Mean (AI)**	**Mean (Human)**	**Difference (Δ)**
Semantic Accuracy	3.54	3.58	+0.04
Terminology	3.33	3.66	+0.33
Technical Detail	3.31	3.57	+0.26

It is worth noting that Δ represents the difference between the mean scores of Human and AI, i.e., Δ = Mean (Human) - Mean (AI). Human translation consistently outperformed AI-assisted translation in all three sub-criteria for accuracy. The highest Δ was in terminology handling, where the Δ was 0.33.

The qualitative results offer direct evidence to explain the results. The interviewees agreed that the most difficult domains for AI translation were medical, legal, technical, and economic terminology. A particularly critical issue that was raised in several interviews was the issue of AI hallucination.

> "*Nó là con dao hai lưỡi — nếu không hiệu chỉnh sẽ dẫn đến sai kết quả và hậu quả khôn lường.*" (It is a double-edged sword --without post-editing, it leads to incorrect results and unpredictable consequences.) (P3)

The participants also mentioned that hallucination risk was heavily affected by the quality of user prompts. Inadequate prompts resulted in syntactically fluent but semantically garbled responses, again underscoring the need to include AI literacy in translation pedagogy.

> "*Phải biết chừng mực trong sử dụng AI, phải phát huy khả năng của con người, xem AI như công cụ hỗ trợ, không thể thay thế hoàn toàn*." (One must know the limits of AI use, must develop human capacity, and treat AI as a support tool that cannot fully replace the translator.) (P7).

The lowest semantic accuracy discrepancy was 0.04, which shows the improvement of AI systems in capturing propositional meaning at the level of sentence. Results showed that Cohesion & Coherence had the highest mean score level for both groups, AI at M = 3.51 and Human at M = 3.64. Moreover, 50.7% of the AI's translations scored Level 4 in Cohesion & Coherence. This is because the transformer model can provide discourse connectedness.

However, the advantage of Human translation is consistent across all sub-criteria for Fluency. The results from the Naturalness sub-criterion (Δ = 0.06) showed that, while AI translations are grammatically correct, they often resulted in mechanical translations lacking the naturalness associated with proficient human translations. A particularly interesting observation is the poor performance by AI with regard to euphemism and socio-emotional expressions specific to Vietnamese discourse.

4.2.2 Fluency

Fluency was evaluated through grammatical correctness, naturalness of expression, and cohesion and coherence. Table 4.6 presents mean scores.

Table 4.6: Mean Scores: Fluency Criteria (AI vs. Human Translation)

Criterion	Mean (AI)	Mean (Human)	Difference (Δ)
Grammatical Correctness	3.33	3.43	+0.10
Naturalness	3.31	3.37	+0.06
Cohesion & Coherence	3.51	3.64	+0.13

For Cohesion and Coherence, the results showed the highest mean values for both groups, i.e., AI: M = 3.51 and Human: M = 3.64. Moreover, 50.7% of the AI output was rated Good, which was the maximum rating AI achieved across all nine dimensions. This suggests that the current NMT systems have improved significantly in retaining the connection and logical order of the text, which is consistent with Koehn's (2020) assessment of NMT's capabilities in sentence-level coherence. However, the human translations outperformed on all dimensions. The naturalness criterion yielded a low Δ value of 0.06, which means AI translation commonly produced mechanically formed Vietnamese sentences.

This was verified by the results of the survey, in which 68.6% of participants agreed that "AI translation sounds mechanical." The interviewees further elaborated that "mechanical output" specifically means coarse phrasing (Vietnamese: lời văn thô), contextually inappropriate word choice, and lack of natural linguistic variation. P5 described this briefly but accurately: "*câu từ khác người viết*" ("the phrasing is different from how a person would write it").

Notably absent in the performance of the AI model were its difficulties with euphemisms (nói giảm nói tránh) and socio-emotional subtleties particular to the Vietnamese language, including the expressed politeness, deference, and tone associated with Vietnamese communicative culture. As such, the participants overwhelmingly perceived the performance of the AI model as being that of an "advanced dictionary/drafting assistant" and not as a cognitively independent translation partner. As such, the participants' performance is consistent with the limitations of NMT technology as discussed by Nurminen *et al.* (2023). Grammatical correctness revealed a moderate gap between the two groups (Δ=0.10), which suggested that although the performance of the AI model was grammatically correct, the participants were more attuned to Vietnamese grammatical structures.

4.2.3 Cultural Appropriateness

Cultural appropriateness was measured through idiomatic expression handling, register, formality, and cultural sensitivity. Table 4.7 presents mean scores.

Table 4.7*:* Mean Scores of Cultural Appropriateness Criteria (AI vs. Human Translation)

Criterion	Mean (AI)	Mean (Human)	Difference (Δ)
Idioms & Metaphors	3.51	3.66	+0.15
Register & Formality	3.31	3.36	+0.05
Cultural Sensitivity	3.51	3.57	+0.06

The cultural appropriateness metric was also a domain in which both groups scored relatively low, but human translators performed better consistently. The discrepancy was bigger for the idiomatic expression measure (Δ = 0.15).

> "*Tiếng Việt khó, phong phú từ ngữ, mang tính chất địa phương nên AI khó có thể nắm bắt.*" (Vietnamese is lexically rich and regionally nuanced in ways AI finds difficult to capture.) [P5]

The cultural sensitivity measure showed a small gap (Δ=0.06), indicating the limited ability of AI tools to deal with the socio-cultural norms inherent in Vietnamese professional and academic writing.

Four patterns of convergence have been found in the results obtained from both quantitative and qualitative methods. First, the biggest gaps have been found for terminology (Δ=0.33) and technical detail preservation (Δ = 0.26), reinforcing the relative weakness of AI systems in terms of accuracy. Second, cultural mediation, specifically idiomatic expression handling, is a second area where human performance is consistently better than AI's. Third, the gaps for semantic accuracy and naturalness have been found to be the smallest for AI systems (Δ=0.04 and Δ=0.06, respectively), suggesting a relative strength of AI systems in terms of grammatical correctness. Fourth, the results obtained from the qualitative methods have shown a pattern where AI systems' fluency may conceal underlying accuracy issues, a pattern consistent with the fluency bias described in Martindale & Carpuat (2018).

The results obtained in this study reinforce the position that AI systems are useful tools for draft generation, producing structurally coherent and semantically adequate translations, but need to be subject to human post-editing for terminological, idiomatic, and cultural appropriateness.

4.2.4 Summary of RQ1 Findings

Table 4.8: Consolidated Comparison: AI vs. Human Translation Across All Criteria

Dimension	Criterion	Mean AI	Mean Human	Δ
Accuracy	Semantic Accuracy	3.54	3.58	0.04
	Terminology	3.33	3.66	0.33
	Technical Detail	3.31	3.57	0.26
Fluency	Grammar	3.33	3.43	0.10
	Naturalness	3.31	3.37	0.06
	Cohesion - Coherence	3.51	3.64	0.13
Cultural	Idioms & Metaphors	3.51	3.66	0.15
	Register & Formality	3.31	3.36	0.05
	Cultural Sensitivity	3.51	3.57	0.06

Four converging patterns can be identified on the basis of quantitative and qualitative findings. Firstly, the performance gap was largest for terminology (Δ=0.33) and technical detail (Δ = 0.26), thus confirming AI's limitations in precision-dependent translation, as directly reflected in interviewees' accounts of systematic "hallucination," "reversal," and the need for post-editing. Secondly, AI's performance was stronger on some dimensions, namely, cohesion and coherence (Δ= 3.51) and semantic accuracy (Δ =3.54), thus implying a level of competence on a surface-structural level. Thirdly, and perhaps most interestingly, no criterion of AI translation ever exceeded a "Very Good" rating, whereas human translation exceeded on all nine dimensions ("Very Good" = 3.0% to 10.4%), thus implying a qualitative performance ceiling for AI translation. Fourthly, register and formality pointed to a shared area of difficulty for both AI and novice human translation, thus implying a priority area for pedagogical intervention, irrespective of technology use.

These results lend support to the view that the tools can be used effectively as draft-generating tools with the capacity to produce acceptable baseline translations, although they require systematic human post-editing, especially with regard to terminology, pragmatic nuances, and cultural adaptation to achieve professionally adequate target texts.

4.3 Research Question 2: Student Perceptions of AI-assisted Translation

RQ2 was also answered through an analysis of the 23-item survey questionnaire, where a reliability coefficient was found to be 0.884, and a total of 67 participants took part in the survey. Four thematic clusters were identified: perceived translation quality, perceived benefits, AI dependency, and pedagogical perceptions. The interviews allowed for a more in the eight-depth interpretation of the quantitative findings.

4.3.1 Perceived Quality and Benefits of AI Translation

The perceptions of participants on the quality of AI translation were found to be rather guarded. Questions related to AI's ability to interpret contextual meanings and idiomatic expressions were found to have scored an average of 2.81-2.94 on a 5-point scale, indicating that students were rather doubtful about AI's ability to grasp the nuances of language.

However, perceptions related to AI's instrumental benefits were found to be more favorable. The highest score was recorded on the question "Do you think AI is a valuable tool for improving translation efficiency?" with a mean score of 3.75. This was also the highest score recorded in the data set.

The in-depth interviews provided more contextual insights into this dual perception of AI translation. The students recognized the benefits of AI translation but were also aware that these benefits were dependent on post-editing and critical evaluation.

4.3.2 AI Dependency

The results of these items regarding dependency on AI showed a moderate level of dependency on AI, as measured by the mean values ranging from 2.93 to 3.36. The item "I struggle to complete translation tasks without AI's help" showed a mean value of 3.21, which meant that one out of every three students acknowledged a certain level of dependency on AI. On the contrary, "I prefer using AI rather than translating independently" showed the lowest mean value in these items, which is 2.93.

Interviewees' perception of dependency on AI showed a certain level of dependency in a task-contingent manner; in other words, they used AI when they were having problems in certain terms or structure, yet they were confident in their own judgment in certain content areas. A more systemic problem related to a cultural shift in a "fast and cheap" translation mentality. Interviewees also warned that if the over-reliance on AI is continued, it can gradually lead to the loss of the sense of translating the translational intuition developed through strategic practice, with the ability to evaluate the translation critically. As O'Brien (2012) states in the case of the post-editing model, the strategic integration of MT is only beneficial to those translators who already possess the level of competence to recognize and correct the errors produced by the AI themselves.

4.3.3 Skill Development and Pedagogical Perception

The participants showed a high sense of pedagogical awareness, with 50.8% agreeing that they need to be guided to use AI effectively in translation, with an average score of 3.51. A positive agreement was found in the aspect of skill development where AI can assist in the development of translation skills with an average score of 3.58 and in the learning of vocabulary and structure with an average score of 3.33.

The results of the interviews showed that there is a three-pillar framework in the use of AI, which includes Skills, or prompt engineering; Mindset, or critical distance from AI; and Attitude, or the value of creative and culturally sensitive expression, which can be used to develop a critically mediated approach to the integration of AI in translation.

From the results of RQ1 and RQ2, it can be deduced that there are obvious efficiency and drafting benefits to AI, but human intervention is necessary to ensure quality, especially in terms of terminological accuracy, cultural mediation, and naturalness. However, students have shown awareness of the potential and limitations of AI, which means they can already benefit from pedagogical intervention, which will be further discussed in the next chapter.

4.3.4 Summary of RQ2 Findings

Table 4.9: Summary of Student Perceptions by Thematic Cluster

Thematic Cluster	Mean Range	Dominant Trend
AI Translation Quality	2.81–2.94	Skeptical — doubts linguistic depth
Perceived Benefits	3.46–3.75	Positive — values efficiency and utility
AI Dependency	2.93–3.36	Moderate — task-contingent reliance
Pedagogical Perception	3.33–3.51	Supportive — wants structured guidance

Five conclusions can be drawn based on the combined results. First, the students demonstrate a pragmatic embrace of AI tools, appreciating the speed and productivity benefits, but not unconditionally trusting the quality of the result. Secondly, the strongest questions regarding the quality of translation provided by the AI tool are associated with its contextual and cultural competency, while the majority agreed that the translation produced by the tool is wrong. Third, the level of dependency is modest which means the students are still dependent on their own translation effort. Fourth, the respondents directly pointed out the potential negative repercussions of the overreliance on the AI tool, such as the loss of cảm giác dịch and pragmatic skill. Finally, the high demand for pedagogical support with the application of the AI tool (50.8%) indicates the importance of the pedagogical role in the integration of the AI tool into the teaching process.

In summary, the results of both RQ1 and RQ2 show the same picture. Although the application of the AI tool can bring benefits to the translation process, it is still not as good as the translation done by humans in all aspects.

5. Conclusion

5.1 Summary of Key Findings

The present study sought to compare the quality of AI-assisted translation and human translation in the context of the English-Vietnamese language pair among third-year students at Binh Duong University, in addition to examining the students' perceptions of AI translation. The results of the present study have consistently revealed three major conclusions.

The first is that human translation has shown superior quality compared to AI translation in all nine criteria. The most significant differences in translation quality were found in terminology use (Δ=0.33) and technical detail (Δ=0.26), which highlighted the shortcomings of AI translation. Notably, the quality of AI translation has been limited to the Fair-to-Good level in all nine criteria, failing to achieve the Very Good level in any of these criteria, while human translation has achieved the Very Good level in all translation dimensions. The second is that AI translation has shown its strength in surface cohesion and semantic accuracy. Third, students hold a nuanced dual perception of AI: they recognize its practical value for efficiency and vocabulary support while simultaneously expressing well-founded skepticism about its cultural depth, contextual accuracy, and reliability in high-stakes translation contexts.

5.2 Pedagogical Implications

The results also hold important implications for the teaching of translation at Binh Duong University and the broader context of higher education in Vietnam. As the use of AI is so integral to the students' work and 50.8% of the students call for instruction on the use of AI, the issue is not whether AI is taught but rather how best to go about teaching it.

One suggestion is that the curriculum include instruction on AI literacy, enabling students to think critically about the product rather than simply accepting the information provided by the AI program. This is in line with Bowker's (2020) suggestion that the teaching of MT needs to move from an approach of avoidance to one of critique. Another suggestion is that post-editing of machine translation (MTPE) is included as an element of the curriculum, with students required to recognize the errors made by the AI program in terms of terminology, register, and cultural appropriateness. This fills the important gap in the post-editing skillset identified in the present study, as well as the professional market situation described by O'Brien (2012). Third, the skill of prompt engineering must be taught. This means that students must be taught to create precise and rich prompts that reduce the incidence of hallucination and steer the AI towards contextually appropriate equivalents. Fourth, comparative AI-HT assignments that ask students to consider the differences between AI and human translation can help develop the kind of translatorial thought required to use AI as an adjunct rather than a crutch. Fifth, the assessment criteria must be revised to reward cultural sensitivity, pragmatic appropriateness, and the post-editing quality aspects that AI tools consistently fail to deliver on, as well as the uniquely human translation skillset.

5.3 Recommended Solutions

Based on the integrated findings, three potential solutions to the issue of translation quality in AI-assisted translation workflows are proposed, informed by the integrated findings, are proposed. Firstly, the application of structured post-editing workflows is recommended as a pedagogical tool to be included in translation courses, in which students learn to assess and improve AI output. Secondly, AI prompting workshops should be designed to assist students in producing high-quality AI output through effective prompting, minimizing the possibility of hallucinations, and improving terminological accuracy. Thirdly, comparative translation tasks should be included in assessment frameworks to help students become aware of the translation process through documenting and justifying their improvements to AI output.

5.4 Limitations of the Study

There are some limitations that should be noted. Firstly, the subjects were limited to third-year English major students in one university; therefore, the results may not generalize well to other types of students and proficiency levels. Secondly, the genre of text used in this study is limited to informative text in academic writing; therefore, it may not generalize well to all genres of translation texts in real-life situations. Thirdly, as the time frame of the study is limited, there is a lack of longitudinal measurement of how the use of AI can influence the development of translation competence in the long term. Lastly, while the results of inter-rater reliability were high at 0.676, they were slightly lower than the optimal level of 0.70; therefore, more extensive training is necessary in future studies.

5.5 Recommendations for Future Research

The limitations identified in the current research should be addressed in future research, and the research should be extended in various ways. Firstly, more participant groups should be included in research conducted in various universities in Vietnam to increase the scope of the research. Secondly, comparative research on different text genres, including literary translation, legal translation, and technical translation, should be conducted to attain a more complete understanding of what AI can and cannot do. Thirdly, longitudinal research on the long-term effects of AI on the competence development of novice translators should be conducted to attain a better understanding of whether AI is a facilitator or inhibitor of novice translators' competencies. Fourthly, research on various strategies of prompt engineering and their effects on the quality of AI translation should be conducted to address the mediating variable identified in the current research. Lastly, comparative research on various pedagogical interventions, including MTPE, error analysis, and AI literacy, should be conducted to inform practitioners on how to design an effective curriculum. Finally, these research outcomes should be utilized to develop appropriate training programs and enhance the quality of translation courses at Binh Duong University, ensuring that students are well-prepared to adapt to the advancements of modern translation technology."

Conflict of Interest Statement
The authors declare no conflicts of interest.

About the Authors
Bùi Kim Phượng is a lecturer at the Informatics & Foreign Language Center, Binh Duong University (IFLC-BDU), Vietnam. She has been teaching English as a Foreign Language (EFL) at the tertiary education level for two years. She can be reached at phuongbinhduongbui@gmail.com.
Nguyễn Minh Trang is a lecturer at the Faculty of Foreign Languages, Binh Duong University, Ho Chi Minh City, Vietnam. He has been teaching English as a Foreign Language (EFL) at the tertiary level for over 25 years. He has several publications in international journals. His research interests include academic listening, ELT issues, translation and interpretation studies. He can be reached at nguyenminhtrang@hotmail.com.

References

Al-Afnan, M. A. (2024). Large Language Models as Computational Linguistics Tools: A Comparative Analysis of ChatGPT and Google Machine Translations. *Journal of Artificial Intelligence Theory*, *5*, 1–15. https://doi.org/10.37965/jait.2024.0549

Arnold, D., Balkan, L., Humphreys, G., Meijer, S., & Robertson, S. (1994). *Machine translation: An introductory guide*. NCC Blackwell.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural machine translation by jointly learning to align and translate*. arXiv. https://arxiv.org/abs/1409.0473

Baker, M., & Saldanha, G. (Eds.). (2020). *Routledge encyclopedia of translation studies* (3rd ed.). Routledge. https://doi.org/10.4324/9781315678626

Bell, R. T. (1991). *Translation and translating: Theory and practice*. Longman. Retrieved from https://books.google.ro/books/about/Translation_and_Translating.html?id=QLxrAAAAIAAJ&redir_esc=y

Bowker, L. (2020). *Machine translation literacy instruction for international students*. John Benjamins. https://doi.org/10.1080/08963568.2020.1794739

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203771587

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications. Retrieved from https://books.google.ro/books/about/Research_Design.html?hl=es&id=335ZDwAAQBAJ&redir_esc=y

Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press. Retrieved from https://books.google.ro/books/about/Research_Methods_in_Applied_Linguistics.html?id=gNh6QgAACAAJ&redir_esc=y

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications. Retrieved from https://books.google.ro/books/about/Discovering_Statistics_Using_IBM_SPSS_St.html?hl=id&id=c0Wk9IuBmAoC&redir_esc=y

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82. https://doi.org/10.1177/1525822X05279903

House, J. (2015). *Translation quality assessment: Past and present*. Routledge. https://doi.org/10.4324/9780203072934

Hutchins, J. (1995). Machine translation: A concise history. *Computing Research Quarterly*. Retrieved from https://doi.org/10.65961/jts-2010-1_2-003

Jakobson, R. (2004). On linguistic aspects of translation. In L. Venuti (Ed.), *The translation studies reader* (2nd ed., pp. 138–143). Routledge. (Original work published 1959). Retrieved from https://web.stanford.edu/~eckert/PDF/jakobson.pdf

Jiao, W., Wang, W., Huang, J. T., Wang, X., & Tu, Z. (2023). Towards making the most of ChatGPT for machine translation. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 6262–6278). Association for Computational Linguistics. https://aclanthology.org/2023.findings-emnlp.373

Koehn, P. (2020). *Neural machine translation*. Cambridge University Press. https://doi.org/10.1017/9781108802302

Koponen, A. (2016). Is machine translation post-editing worth the effort? A usability study with student translators. *JoSTrans: The Journal of Specialised Translation*, *25*.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lörscher, W. (1991). *Translation performance, translation process and translation strategies*. Retrieved from https://books.google.ro/books/about/Translation_Performance_Translation_Proc.html?id=aanlAAAAMAAJ&redir_esc=y

Mammadova, I. (2024). Machine translation vs. human translation: A comparative analysis. *Journal of Language and Translation Studies*, *12*(2), 45–67. Retrieved from https://egarp.lt/index.php/JPURM/article/view/174

Martindale, M. J., & Carpuat, M. (2018). Fluency over adequacy: A pilot study in measuring user trust in imperfect MT. *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas*, 13–25. https://aclanthology.org/W18-6302/

Nababan, M. R., Nuraeni, A., & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. *Kajian Linguistik và Sastra*, *17*(1), 39–57.

Nguyen, K. M., Nguyen, T. H., Luu, K. Q., & Nguyen, L. M. (2022). A Vietnamese-English neural machine translation system. In *Interspeech 2022* (pp. 1234–1238). ISCA. https://www.isca-speech.org/archive/interspeech_2022/

Nida, E. A. (1964). *Toward a science of translating*. Brill. Retrieved from https://books.google.ro/books/about/Toward_a_Science_of_Translating.html?id=T9A7EAAAQBAJ&redir_esc=y

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill. Retrieved from https://books.google.ro/books/about/Psychometric_Theory.html?id=WE59AAAAMAAJ&redir_esc=y

O'Brien, S. (2012). Translation as human-computer interaction. *Translation Spaces*, *1*(1), 101–122. https://doi.org/10.1075/ts.1.06obr

OECD. (2022). *AI and the future of skills, Volume 1: Capabilities and assessments*. OECD Publishing. https://doi.org/10.1787/5b324866-en

PACTE Group. (2011). Strategy and translation competence. In K. Windle & A. Pym (Eds.), *The Oxford handbook of translation studies* (pp. 391–402). Oxford University Press.

Pym, A. (2013). Translation skill-sets in a machine-translation age. *Meta*, *58*(3), 487–503. https://doi.org/10.7202/1025047ar

Reiss, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. Hueber. Retrieved from https://books.google.ro/books/about/M%C3%B6glichkeiten_und_Grenzen_der_%C3%9Cbersetz.html?id=M31JAQAAIAAJ&redir_esc=y

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson. Retrieved from https://books.google.ro/books/about/Using_Multivariate_Statistics.html?id=ucj1ygAACAAJ&redir_esc=y

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381115

Vaswani, A., *et al.* (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 30. https://arxiv.org/abs/1706.03762

Vermeer, H. J. (1989). Skopos and commission in translational action. In A. Chesterman (Ed.), *Readings in translation theory* (pp. 173–187). Finnlectura. Retrieved from https://www.taylorfrancis.com/chapters/edit/10.4324/9780429280641-26/skopos-commission-translational-action-hans-vermeer-andrew-chesterman

Wei, J., *et al.* (2025). *Representation drift in long-context machine translation*. arXiv preprint arXiv:2505.16894.

Williams, J. (2013). *Theories of translation*. Palgrave Macmillan. https://books.google.ro/books/about/Theories_of_Translation.html?id=-agox8ZrSg8C&redir_esc=y

Wolff, J. (2024). Zero-shot learning and its implications for neural machine translation. *AI & Society*, *39*(1), 112–125.

Zouhar, V., Tamchyna, A., Popel, M., & Bojar, O. (2021). Neural Machine Translation Quality and Post-Editing Performance. *In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 10204–10214). https://doi.org/10.18653/v1/2021.emnlp-main.801

Appendices

Appendix 1: Survey Questionnaire Sample

1 = Strongly disagree (Hoàn toàn không đồng ý) 2 = Disagree (Không đồng ý)
3 = Neutral (Trung lập) 4 = Agree (Đồng ý) 5 = Strongly agree (Hoàn toàn đồng ý)

Part I: Background Information/Thông tin cá nhân

1. Gender (Giới tính)
☐ Male (nam) ☐ Female (nữ)
2. Major (ngành học) ☐ English (tiếng Anh) ☐ Non-English (ngành khác)
3. Year of study (Năm học tập)
☐ First year ☐ Second year ☐ Third year ☐ Fourth year
4. Frequency of AI use in your English translation
☐ Never (chưa bao giờ) ☐ Rarely (hiếm khi) ☐ Sometimes (thỉnh thoảng)
☐ Often (thường xuyên) ☐ Very often (rất thường xuyên)
4. Which AI tools do you use for English–Vietnamese translation?
☐ Google Translate ☐ DeepL ☐ ChatGPT/Claude/Gemini
☐ Other: ____
5. What is your main purpose for using AI in English–Vietnamese translation?
☐ Looking up vocabulary/terminology
☐ Producing a rough translation of a whole paragraph
☐ Checking grammar/style
☐ Suggesting different translation options
☐ Other: ____

Part II: Comparison between AI and Human Translation/So sánh dịch sử dụng AI và dịch con người.

I. Accuracy/Độ chính xác

1. Do you think AI-assisted translation is as accurate as human translation? (Bạn có cho rằng dịch có hỗ trợ AI chính xác ngang với dịch của con người không?)
2. Do you think AI correctly translates specialized terminology? (Bạn có cho rằng AI dịch chính xác các thuật ngữ chuyên ngành không?)
3. Do you think human translation handles ambiguous meanings better than AI? (Bạn có cho rằng dịch của con người xử lý nghĩa mơ hồ tốt hơn AI không?)
4. Do you think AI sometimes produces contextually incorrect translations? (Bạn có cho rằng AI đôi khi tạo ra bản dịch sai về mặt ngữ cảnh không?)

II. Fluency/Độ trôi chảy

5. Do you think AI translations are natural and fluent in Vietnamese? (Bạn có cho rằng bản dịch của AI tự nhiên và trôi chảy bằng tiếng Việt không?)
6. Do you think AI-generated texts sometimes sound mechanical? (Bạn có cho rằng văn bản do AI tạo ra đôi khi nghe máy móc không?)

7. Do you think human translations are more coherent than AI translations?(Bạn có cho rằng bản dịch của con người mạch lạc hơn bản dịch của AI không?)
8. Do you think AI produces grammatically correct sentences?(Bạn có cho rằng AI tạo ra các câu đúng ngữ pháp không?)

III. Cultural Appropriateness/Tính phù hợp văn hoá

9. Do you think AI can appropriately translate culture-specific expressions? (Bạn có cho rằng AI có thể dịch phù hợp các biểu đạt mang tính đặc thù văn hoá không?)
10. Do you think AI fails to convey cultural nuances effectively? (Bạn có cho rằng AI không truyền tải hiệu quả sắc thái văn hoá không?)
11. Do you think human translators adapt translations better to Vietnamese culture? (Bạn có cho rằng dịch giả con người điều chỉnh bản dịch phù hợp với văn hoá Việt Nam tốt hơn không?)
12. Do you think AI can successfully translate idioms and figurative language? (Bạn có cho rằng AI có thể dịch thành công thành ngữ và ngôn ngữ ẩn dụ không?)

Part III: Third-Year Students' Perception of AI/Nhận thức của sinh viên năm 3 về AI.

Part IV: Perceived Usefulness/Nhận thức về tính hữu ích.

13. Do you think AI improves the overall quality of your translations? (Bạn có cho rằng AI cải thiện chất lượng tổng thể bản dịch của bạn không?)
14. Do you think AI helps you complete translation tasks more efficiently? (Bạn có cho rằng AI giúp hoàn thành nhiệm vụ dịch hiệu quả hơn không?)
15. Do you think AI enhances your productivity in translation assignments? (Bạn có cho rằng AI nâng cao năng suất làm bài tập dịch không?)
16. Do you think AI is a valuable tool in translation learning? (Bạn có cho rằng AI là một công cụ có giá trị trong việc học dịch không?)

Part V. AI Dependency/Mức độ phụ thuộc vào AI.

17. Do you rely on AI to generate initial drafts? (Bạn có phụ thuộc vào AI để tạo bản nháp ban đầu không?)
18. Do you feel less confident translating without AI? (Bạn có cảm thấy kém tự tin hơn khi dịch mà không có AI không?)
19. Do you think you would struggle to complete translation tasks without AI? (Bạn có nghĩ rằng sẽ gặp khó khăn khi hoàn thành bài dịch nếu không có AI không?)
20. Do you prefer using AI even for simple translation tasks? (Bạn có thích sử dụng AI với các nhiệm vụ dịch đơn giản không?)

Part VI. Skill Development & Pedagogical Perception/Phát triển kỹ năng & Nhận thức sư phạm.

21. Do you think AI helps you understand translation strategies better? (Bạn có cho rằng AI giúp hiểu chiến lược dịch tốt hơn không?)

22. Do you think using AI improves your translation competence over time? (Bạn có cho rằng việc sử dụng AI giúp nâng cao năng lực dịch theo thời gian không?)
23. Do you think AI should be integrated into translation courses? (Bạn có cho rằng nên tích hợp AI vào các học phần dịch thuật không?)
24. Do you think lecturers should provide guidance on how to use AI effectively? (Bạn có cho rằng giảng viên nên hướng dẫn cách sử dụng AI hiệu quả không?)

Part VII. Open-ended Questions

The following questions aim to collect additional qualitative insights to complement the survey results."(Phần câu hỏi dưới đây nhằm thu thập thêm dữ liệu định tính bổ trợ cho kết quả khảo sát.)

25. In your opinion, what is the biggest advantage of using AI in translation? Why? (Theo bạn, lợi ích lớn nhất của việc sử dụng AI trong dịch thuật là gì? Vì sao?)
26. What is the biggest limitation of AI in translation, and how can it be improved? (Hạn chế lớn nhất của AI trong dịch thuật là gì và có thể cải thiện như thế nào?)

Appendix 2: Sample Translation Text

(Translate into Vietnamese by two translation versions: The first version without AI, and the second one with AI.)

In recent years, artificial intelligence has significantly transformed the translation industry in unimaginable ways. AI-assisted translation tools are now capable of generating drafts within seconds, offering remarkable efficiency and convenience. However, such technological advancement is not always a blessing in disguise. While AI can process large volumes of text quickly, it often struggles with specialized terminology and context-sensitive expressions. Terms such as "pragmatic competence," "discourse analysis," and "contextual inference" require interpretive precision rather than literal, word-for-word rendering. A mechanical translation of these concepts may distort their academic meaning, particularly when ambiguity or cultural nuance is involved. Furthermore, English academic writing typically favors directness and explicit argumentation, whereas Vietnamese discourse often emphasizes subtlety, indirectness, and relational harmony. This stylistic divergence requires translators to carefully adjust tone and structure to suit the target audience. Without thoughtful post-editing, AI-generated translations may appear grammatically accurate yet stylistically rigid and culturally detached. Ultimately, effective translation extends beyond lexical equivalence; it involves conveying meaning, tone, and cultural nuance in a manner that resonates authentically with readers in the target language community.

1. *The first translation version without AI (Người dich không có sự hỗ trợ của AI)*

……
……
……

2.The second translation version with AI (Người dịch có sự hỗ trợ của AI)

……
……
……

Appendix 3: Semi-Structured Interview

Topic: Benefits and Limitations of AI in Translation

Part 1: General Information/Thông tin chung

Interviewee (Người được phỏng vấn) .....
Role/Occupation (Vai trò/Nghề nghiệp): .....
Date (Ngày phỏng vấn): .....

Part 2: Interview Contents/Nội dung phỏng vấn

I. Accuracy & Terminology/Độ chính xác & Thuật ngữ

1. "Based on your class assignments, which specific fields of terminology (e.g., medical, legal, or idioms) do you find AI struggling with the most because it fails to understand the context?) (Dựa trên các bài tập trên lớp, bạn thấy AI gặp khó khăn nhất với loại thuật ngữ chuyên ngành cụ thể nào (ví dụ: y tế, pháp lý, hay thành ngữ) do nó không hiểu được ngữ cảnh?)
2. Have you experienced AI "hallucination" or meaning reversal? How did you handle it? (Bạn đã gặp trường hợp AI "tự bịa" (hallucination) hoặc dịch ngược nghĩa chưa? Bạn xử lý thế nào?)
3. In your opinion, does accuracy depend on the AI model itself or the user's prompting? (Theo bạn, độ chính xác phụ thuộc vào bản chất mô hình AI hay cách người dùng đặt câu lệnh (prompt)?)

II. Fluency & Style/Sự trôi chảy và văn phong

4. How do you define a "mechanical" translation? What are the signs of an AI-generated text? (Bạn định nghĩa thế nào là bản dịch "máy móc"? Dấu hiệu nào giúp nhận ra đó là sản phẩm của AI?)
5.Which types of texts (literature, ads, or technical reports) are most difficult for AI to translate fluently? (AI gặp khó khăn nhất với loại văn bản nào (văn chương, quảng cáo, hay báo cáo kỹ thuật) để trôi chảy?)
6."What specific post-editing steps do you take to fix AI errors and make the sentences sound more natural?" (Bạn thực hiện các bước hậu kiểm cụ thể nào để sửa lỗi và giúp câu văn tự nhiên hơn?)

III. Cultural Appropriateness/Sự phù hợp về văn hóa

7.Why is it difficult for AI to grasp Vietnamese cultural nuances or metaphors (localisms, euphemisms)? (Tại sao AI khó nắm bắt sắc thái văn hóa hoặc ẩn dụ trong tiếng Việt (từ địa phương, nói giảm nói tránh)?
8. Do you believe AI will understand "emotions" and "social contexts" in the future? Why or why not? (Bạn có tin tương lai AI sẽ hiểu được "tình cảm" và "ngữ cảnh xã hội" không? Tại sao? Tại sao không?)

9. When translating texts that require high localization or emotional nuance, do you see AI as a 'collaborative assistant' or just a 'glorified dictionary'? (Khi dịch các văn bản đòi hỏi tính bản địa hóa cao hoặc sự tinh tế về cảm xúc, bạn xem AI là một 'cộng sự hỗ trợ' hay chỉ là một 'cuốn từ điển nâng cao'?)

IV. Benefits & Limitations/Lợi ích & Hạn chế

10. Besides saving time, does AI help you expand your vocabulary or suggest new ways of expression? (Ngoài tiết kiệm thời gian, AI có giúp bạn mở rộng vốn từ hoặc gợi ý cách diễn đạt mới không?)

11. Do you worry that using AI will make you lose your 'translation sense' or critical thinking?" (Bạn có lo lắng việc dùng AI sẽ làm bạn mất đi 'cảm giác dịch' hoặc tư duy phản biện không)

12. Is the current trend leaning towards "Fast-cheap (AI)" or "Slow-quality (Human)" translation? Hiện nay xu hướng đang nghiêng về "Dịch nhanh-rẻ (AI)" hay "Dịch chậm-chất lượng (Người))"?

Part 3: Closing/Kết thúc

Additional Question/Câu hỏi bổ sung

Is there anything else you would like to share about this topic? (Anh/chị còn điều gì muốn chia sẻ thêm về vấn đề này không?)

Appendix 4: Translation Quality Scoring Rubric (1–5 Linkert Scale)

Score	Accuracy	Fluency	Cultural Appropriateness
5 (Excellent)	All key ideas accurately conveyed; specialized terms precisely translated; no distortion.	Natural, coherent, stylistically appropriate Vietnamese; smooth flow; no awkward structures.	Idioms effectively adapted; tone fully appropriate for Vietnamese academic discourse.
4 (Good)	Minor lexical inaccuracies; terminology mostly correct; meaning largely preserved.	Generally fluent; minor awkward phrasing; coherence mostly maintained.	Minor cultural mismatches; tone generally appropriate.
3 (Fair)	Some meaning shifts, partial mistranslation of terms, and minor omissions.	Noticeable unnatural expressions; some mechanical structures.	Literal translation of idioms, but understandable; limited cultural adaptation.
2 (Weak)	Significant inaccuracies; several terminology errors; meaning partially distorted.	Frequent awkward or literal translation; poor cohesion; mechanical tone.	Cultural expressions poorly handled; inappropriate tone or register.
1 Poor)	Major meaning distortion; key ideas mistranslated or omitted; terminology incorrect.	Very unnatural; difficult to read; fragmented or incoherent.	Serious cultural misunderstanding; idioms mistranslated; tone unsuitable.