

Investigating the Integration of LLMs into Trainee Translators' Practice and Learning: A Questionnaire-based Study on Translator-AI Interaction

Xindi Hao & Shuyin Zhang*

The Chinese University of Hong Kong, Shenzhen, Guangdong, 518172, P.R. China

zhangshuyin@cuhk.edu.cn

Abstract

In recent years, large language models (LLMs) have drawn significant attention from translators, including trainee translators, who are increasingly adopting LLMs in their translation practice and learning. Despite this growing interest, to the best of our knowledge, no LLM has yet been specifically designed for (trainee) translators. While numerous LLMs are available on the market, their potential in performing translation-related tasks is yet to be fully discovered. This highlights a pressing need for a tailored LLM translator guide, conceptualized as an aggregator or directory of multiple LLMs and designed to support trainee translators in selecting and navigating the most suitable models for different scenarios in their translation tasks. As an initial step towards the development of such a guide, this study aims to identify the scenarios in which trainee translators regularly use LLMs. It employs questionnaire-based research to examine the frequency of LLM usage by trainee translators, the average number of prompts, and their satisfaction with the performance of LLMs across the various scenarios identified. The findings give an insight into when and where trainee translators might integrate LLMs into their workflows, identify the limitations of current LLMs in assisting translators' work, and shed light on a future design for an LLM translator guide.

1 Introduction

Large language models (LLMs) function as the foundation models of Generative AI (GenAI) in performing text generation and language

processing (Bhupathi, 2025). Very recently, the advent of LLMs has significantly impacted the translation industry. LLMs such as GPT-4, one of the latest in the Generative Pre-trained Transformer (GPT) series, BERT, and LLaMA have quickly become popular tools in translators' workstations, reshaping established practices. In translation industry, there are also translation-specific LLMs or LLM-integrated computer-assisted translation (CAT) tools, such as Trados Copilot and Wordscope, that are primarily designed for translation providers and professional translators. These AI-powered commercial tools provide professional translators with an all-in-one solution for their translation practice (Wordscope). Unlike traditional NMT which is purely an approach to automatic machine translation (Mohamed et al., 2021), with their "inherent ability to understand, generate, and manipulate human-like text in a contextually relevant manner" (Naveed et al., 2023), LLMs can be applied to a wide range of natural language processing (NLP) tasks, including question answering, summarization, text generation, and others. In other words, beyond their direct application to translation in the narrow sense, the high versatility of LLMs and their ability to be customized through prompt engineering can enable them to assist with various tasks across the entire translation workflow.

The potential of LLMs in the translation industry warrants further exploration. In modern translation services, a translation project can, by and large, be divided into three phases: pre-production, production, and post-production, as outlined in the two standards, ISO 17100:2015 and ISO 11669:2024. While these standards are designed to provide guidance for translation service providers from a project management perspective, covering various administrative activities, many of the outlined tasks are also performed by, or involve, individual translators, even during the pre- and

© 2025 The authors. This article is licensed under a Creative Commons 4.0 licence, no derivative works, attribution, CC-BY-ND.

post-production stages. The key stages and tasks a translator may encounter throughout the entire process are summarized in Table 1, adapted from these standards, with tasks more closely aligned

Phase	Tasks / Stages
Pre-production	Setting up translation memories, terminological databases, style-guides
	Preparation of the content for translation technology processing
	Source language content analysis
	Collection and preparation of reference materials
Production	Translation
	Check
	Revision
	Review
	Proofreading
	Final verification and release
Post-production	Feedback collection

Table 1: Three Phases of a Translation Project

with managerial responsibilities excluded, as these typically fall under the role of a project manager.

It is not difficult to envisage LLMs being incorporated into many of these tasks or stages. When examining the task of translation in the narrow sense, it involves several functions where LLMs might be helpful, serving, for instance, as a dictionary, as an machine translation system providing reference translations, or even as a subject-matter expert by offering domain-specific knowledge, not to mention the fact that they could potentially be applied to more complex pre-production tasks, such as content analysis and terminology extraction, as well as in post-production, where they might support feedback collection through the analysis of reviewer comments or client input.

To date, much attention has been directed to claims of human parity in the translation abilities of LLMs, with a particular focus on their performance as machine translation systems—both in terms of evaluation (Hendy et al., 2023) and improvement (Bawden & Yvon, 2023; Moslem et al., 2023). However, scant attention has been paid to the way in which translators, especially trainee translators, integrate LLMs into their daily workflows in practical terms. So far, Sahari et al. (2023) have conducted a cross-sectional study exploring

attitudes of translation teachers and language-related major students towards ChatGPT and Google Translate, and the advantages and challenges brought by ChatGPT. The results show that among four language-related majors, all translation students prefer Google Translate over ChatGPT. Another study conducted by Zhang et al. (2025) investigates how translation students understand the benefits and challenges of using GenAI into their translation practice. While the study examined the functions of GenAI tools used by students in their translation practices, such as looking for background information, generating machine translation outputs, polishing human translations, and providing references for terminologies, its primary purpose was to explore trainee translators' perceptions of using GenAI in translation. However, the actual integration of LLMs into trainee translators' learning and practice remains underexplored. Therefore, this paper addresses this gap by investigating the use of LLMs across the three phases of translation services and their broader impact on human-AI communication with a focus on trainee translators.

To this end, the study examines when, in what contexts, and for what purposes trainee translators incorporate LLMs into their workflows, assessing their effectiveness and efficiency in different translation-related scenarios from a user-centered perspective. The study aims to identify the scenarios in which LLMs are most suitable and effective in students' translation workflow through a survey-based study. The results will serve as the initial step toward developing a large project: the design of an LLM translator guide to help trainee translators choose the most suitable LLM from among numerous options, including scenario-specific LLMs trained for different translation tasks and equipped with preset prompts. With its emphasis on translators in training, this research also seeks to contribute to the development of educational programs to better prepare future professionals for an AI-driven translation industry.

2 Literature Review

Apart from the technically oriented research mentioned above, current scholarly work in translation studies focusing on translators and users mainly addresses the perception and reception of new technologies, particularly AI, by translators (Wang et al., 2024; Wang & Zhang, 2024) and their impact on the language services industry

(Moorkens & Arenas, 2024; Shormani, 2024). More recently, a growing body of literature has begun to examine ethical concerns and sociotechnical effects associated with these innovations (Martinez Carrasco et al., 2024; Moorkens et al., 2024; O'Brien, 2024; Yu & Guo, 2024).

The pedagogical applications and implications of GenAI have also begun to attract considerable attention, particularly in the context of computer-assisted translator training (Ghosh & Chatterjee, 2024; Venkatesan, 2023). For example, Pym and Yu (2024) discuss the way in which translation technologies, including GenAI, can be integrated into language learning and translator training. Similarly, Peng et al. (2024) dedicates an entire section to pedagogy, including insights into students' experiences with, and feedback on, the use of translation technology.

Nevertheless, research on human-AI interaction and the comprehensive application of LLMs throughout all three phases of translation services—pre-production, production, and post-production—remains limited. While certain studies have examined prompting LLMs for translation tasks (Pourkamali & Ebrahim Sharifi, 2024; Zhang et al., 2023), the potential of LLMs to support functions beyond linguistic transfer through prompt engineering has received but little attention. Yamada (2023) investigates ChatGPT's customizability, for instance, but limits its analysis to prompt engineering for enhancing translation quality, so that further research is needed to examine its broader applications within a translator's workstation.

3 LLM-Activated Scenarios

3.1 Constitution of a translator's workstation

Since the concept of the translator's workstation emerged in the 1960s, numerous scholars, beginning with Martin Kay (1980), have attempted to define the range of facilities it might encompass. Among the key contributions to this discourse, Melby (1992) identifies three levels of functions for a translator workstation: (1) word processing, telecommunications and terminology management; (2) text analysis, dictionary lookup, and bilingual text retrieval; and (3) an interface to machine translation systems (147). More recently, Alonso and Nunes Vieira (2017) have updated Kay's (1980)

seminal idea of a translator's amanuensis by proposing the Translator's Amanuensis 2020, which serves both "the general public in their daily translating needs, providing instant machine translation (henceforth referred to as 'the utility level'), and different actors involved with translation in professional settings" (349). Specifically, TA2020 (Alonso & Nunes Vieira, 2017) incorporates the following abilities:

- a) parse the source content (whether written or audio-visual);
- b) identify keywords (key concepts), topics, and genre;
- c) mine virtual content (publicly available and private knowledge bases) and social media in order to find relevant and reliable sources of information to be consulted in the translating process (websites, parallel multilingual content, images, augmented reality output, videos, news, reports), previous translations, and relevant multimodal content. (351)

Ideally, as a critical component of a modern translator's workstation, LLMs should be capable of performing many of these functions while addressing both source- and target-language perspectives.

3.2 Formulation of LLM-activated scenarios

In the preliminary stage of the study, we hypothesized that it was the "chatbot" function of LLMs that would be active when performing translation tasks, particularly their multi-turn dialogue capabilities (Bang et al., 2023). Translation is a decision-making process involving "a series of a certain number of consecutive situations imposing on the translator the necessity of choosing among a certain (and very often exactly definable) number of alternatives," as Levý points out (1967, p. 1171). In this sense, whenever a translator needs to come to a decision, LLMs can provide contextually relevant suggestions, thereby greatly expanding the scope of its application and utility.

Moreover, the real-time interactive query function allows LLMs to answer questions, resembling the search/query function of an internet browser. This means that whenever a translator seeks information, he or she would be able to apply directly to an LLM for assistance. These information retrieval and feedback-seeking functions are the most important ones throughout the process.

To better identify the specific steps or scenarios involved in translation, we drew on the states and events framework proposed by Hlebec (1989) as a reference and adapted it for the purpose of this study.

1. (Activating) knowledge required for an interpretation of the original
2. Choosing the code
3. Interpreting the original
4. Deciding on, or recognizing the presence or absence of the original
5. Considering the form of the translation code
6. Deciding the degree of literalness
7. Determining the intentions
8. Deciding on the manner of conveying the original intentions
9. (Activating) the knowledge required for recoding

In addition, we included the following necessary tasks—checking and revision—as highlighted in ISO 17100:2015, which a translation service might require before the submission and release of a translation, as well as feedback after submission:

10. Checking the target content for semantic, grammatical, and spelling issues, as well as omissions and other errors
11. Examining the target language content against the source language content for any errors, for suitability purpose, and for making corrections
12. Client feedback and satisfaction assessment

To prepare the design of our survey study, we further elaborated on these 12 scenarios and concretized them with a detailed list of functions for LLMs, inspired by Siu (2023):

1. Providing summaries of source texts
2. Highlighting key terms or phrases that require special attention
3. Offering background knowledge or explanations for culturally specific references
4. Suggesting appropriate translations for domain-specific terms
5. Retrieving definitions and usage examples from bilingual corpora or glossaries
6. Automatically identifying inconsistencies in terminology across the text
7. Deciding between literal and free translation based on the purpose of the text
8. Choosing appropriate style, tone, and register for the target audience
9. Resolving ambiguities in the source text

10. Answering specific questions on terminology, grammar, or cultural references
11. Providing links to relevant external resources
12. Acting as an advanced search engine
13. Identifying and correcting grammatical, semantic, or stylistic issues in the target text
14. Comparing the translation with the source text to ensure fidelity and alignment
15. Assessing the target text's suitability for its intended purpose and audience
16. Simulating a client to provide feedback
17. Analyzing client feedback to identify recurring issues or preferences
18. Providing suggestions for future

Phase	Scenarios
Pre-production	Summarizing the content of the source text
	Highlighting key terms or phrases that require special attention
	Providing background knowledge or external resources for understanding the source text
Production	Answering specific questions about terminology, grammar, or cultural references
	Suggesting appropriate style, tone, and register for the translation
	Providing translation references for sentences or paragraphs
	Identifying (and correcting) grammatical, semantic, or stylistic issues in the target text
	Examining whether the translation meets the standard of classic translation norms like “faithfulness, expressiveness and elegance”
Post-production	Providing feedback from the target audience's perspective
	Providing suggestions for future translations based on past feedback

Table 2: Ten Scenarios where Trainee Translators might Use LLMs Throughout the Translation Process

translations based on past feedback

4 Methodology

4.1 Design of the survey

A questionnaire was designed for the purposes of this study in order to investigate the way in which trainee translators use LLMs during a translation task (a task serving the same function for trainee translators as a translation service does for profession translators) by examining three aspects: the frequency of using LLMs in different scenarios, the prompting times in each scenario, and their satisfaction with the performance of LLMs in these scenarios (see Appendix A). The frequency of their LLM usage is used to identify situations where trainee translators commonly use LLMs during the translation process. The prompting times are expected to indicate the extent to which trainee translators strive to interact with LLMs and the efficiency of LLMs when used for different purposes, as fewer rounds of interaction improve user experience. In this context, prompting times refer to the average number of prompts given to the LLM to achieve a specific goal. For example, a trainee translator may prompt an LLM five times to search the background information on a culture-specific term or prompt an LLM three times to check the accuracy of a translation. The effectiveness and suitability of current LLMs under different circumstances is surveyed in “translators’ satisfaction with LLMs”.

The 18 scenarios introduced in Section 3.2, which aim to cover every possible situation where translators might resort to LLMs for a translation task, were further categorized into pre-production, production, and post-production scenarios, based on the phases and stages described in Table 1. Some overlapping scenarios have been streamlined and modified to ensure clearer distinctions and enhance the understanding of participants. As a result, we produced a table of ten refined scenarios (see Table 2).

In the research for the questionnaire, for each scenario, participants were first asked to specify the frequency of their LLM usage and were provided with four options: “never or rarely”, “sometimes”, “often”, and “always”. When participants chose the latter of the three options, which implied that they had access to LLMs and used them for a certain purpose, they would be further asked about their interaction times with LLMs on average and to rate

Phase	Scenarios	Metrics		
		Frequency Score	Prompting Times	Satisfaction Score
Pre-production	Summarizing the content of the source text	5	9	8
	Highlighting key terms or phrases that require special attention	7	10	5
	Providing background knowledge or external resources for understanding the source text	2	7	2
Production	Answering specific questions about terminology, grammar, or cultural references	1	8	1
	Suggesting appropriate style, tone, and register for the translation	6	2	6
	Providing translation references for sentences or paragraphs	4	6	10
	Identifying (and correcting) grammatical, semantic, or stylistic issues in the target text	3	4	4
	Examining whether the translation meets the standard of classic translation norms like “faithfulness, expressiveness and elegance”	8	1	9
Post-production	Providing feedback from the target audience’s perspective	10	5	3
	Providing suggestions for future translations based on past feedback	9	3	7

Table 3: Rankings of the Ten Scenarios Based on the Three Metrics

their performance. However, for those who had “never or rarely” used LLMs in a certain scenario, the questionnaire offered options which were carefully designed to capture the possible reasons, including “I have never thought of using LLMs in this way”, indicating their lack of understanding of

the use of LLMs or awareness of this step during their translation practice, “I think LLMs’ answers are not reliable”, implying their distrust of LLMs, and “I think LLMs’ answers are not useful”, denoting the technical limitations of LLMs. If none of the options was suitable, participants were also asked to write down other underlying reasons. In addition, at the end of the questionnaire, participants were invited to list their most commonly used LLMs.

4.2 Participants

In this study, all the participants were first-year MA students enrolled in the Translation and Interpreting Studies program (with either a Translation and Interpreting major or a Translation plus New Technologies major) or the Simultaneous Interpreting program at the university where the researchers of this paper currently work. All participants were native Chinese speakers whose working language pair was Chinese and English, with IELTS scores of at least 7, who had completed at least one translation-related course during their postgraduate studies and had experience working on translation tasks both individually and in groups. All participants had been introduced to LLMs by their instructors, were familiar with LLMs, and had prior experience of using them in their translation.

4.3 Procedures

The questionnaire content was first submitted to the Applied Psychology Institutional Review Board of the university for an ethical check. Following approval, questionnaires with detailed instructions were distributed to participants via WJX.CN¹, an electronic survey platform widely recognized in China. A total of 50 questionnaires were collected, of which 41 were deemed valid for research purposes.

4.4 Data processing

To investigate the using frequency of each scenario, the interactions with LLMs, and the participants’ evaluation of the performance of LLMs, the study employed a weighted average approach to calculate three metrics: frequency scores, prompting times, and satisfaction scores. For the frequency scores, participants’ responses were weighted as follows: 0 point for “Never or Rarely,” 1 point for “Sometimes,” 2 points for “Often,” and 3 points for

“Always.” To rank the prompting times, we assigned 1 point for the option “1-5,” 2 points for “6-10,” 3 points for “11-15,” 4 points for “16-20,” and 5 points for “Over 20.” For the satisfaction scores, participants rated their satisfaction on a 5-point Likert scale (with 5 representing the highest level of LLM performance). Weighted averages were calculated for all three metrics across the ten scenarios (see Appendix B), and the scenarios were ranked in descending order to identify the most frequently used scenarios, the highest prompting times, and the highest satisfaction scores. Table 3 exhibits the rankings of the ten scenarios for the three metrics.

To better understand the performance of LLMs in each scenario, the researchers calculated the average number of prompts and average satisfaction score of the ten scenarios, then compared the prompting times and satisfaction score of each scenario with the corresponding averages. If the prompting times of a scenario was higher than the average, it may suggest that more time and energy were invested in these scenarios, indicating low efficiency in LLM performance. Conversely, if the prompting times of a scenario was lower than the average, it could mean less efforts spent on that scenario and more efficient LLM performance. Similarly, if the satisfaction score of a scenario was higher than the average, it may suggest participants’ satisfaction with LLMs’ performance in this scenario. However, if the satisfaction score of a scenario was lower than the average, it could mean the unsatisfactory performance of LLMs in this scenario.

5 Analysis

5.1 The interrelationship among the metrics

Given the primary goal of the study—to explore trainee translators’ use of LLMs in their translation workflow, the analysis started from categorizing scenarios into two types, those where translation students regularly used LLMs and those where they rarely did, based on the ranking of the frequency scores. Then, the study examined the prompting times and satisfaction scores of each scenario to understand their popularity, as these two metrics respectively reflected the efficiency and effectiveness of LLM use. For instance, a high satisfaction score of a scenario may explain the

¹ <https://www.wjx.cn/>

frequent use of LLMs under this circumstance, while a high prompting count may suggest more effort was required to prompt the LLM to achieve the goal in this scenario and thus indicate a less satisfactory evaluation and less frequent use.

5.2 Seven regular scenarios where trainee translators use LLMs

In the ranking of the ten scenarios based on the frequency scores (see Table 3, Column Frequency Score), the top seven were recognized by over 50 percent of participants as regular scenarios where they used LLMs (i.e., participants selected “Sometimes,” “Often,” or “Always” as their response). It should be noted that all seven scenarios belonged to pre-production and production stages, indicating that current LLMs were generally more suitable and useful in these phases from the perspective of trainee translators. For these seven regular scenarios, four distinct roles played by LLMs can be observed (see Table 4). In other words, LLMs, like trainee translators’ assistants, are capable of taking on the four specific roles in their translation workflows.

Roles	Functions
Corrector	To proofread trainee translators’ work and to identify grammatical, semantic, or stylistic issues
Explainer	To explain various aspects for trainee translators, including terminology, background knowledge, and register of the source text
Generator	To generate new content by offering translations for certain sentences or paragraphs
Summarizer	To read information, extract key points and summarize the content. Examples include highlighting critical parts that need special attention during translation or summarizing the content of the source text

Table 4: LLMs’ Four Roles in Translation Tasks

The comparison between satisfaction score on average and that of each scenario showed that, four out of the seven regular scenarios—shaded in Table 3, Column Satisfaction Score—namely “answering specific questions about terminology, grammar, or cultural references”, “providing background knowledge or external resources for understanding the source text”, “identifying (and

correcting) grammatical, semantic, or stylistic issues in the target text”, and “highlighting key terms or phrases that require special attention”, scored above average, which, aligns with their frequent use by trainee translators and, to some extent, explains why these functions were frequently used by trainee translators. Trainee translators were satisfied with LLMs’ performance in these scenarios, which belong to the three roles—Corrector, Explainer, and Summarizer. However, though the researcher had assumed that students’ low level of satisfaction with a certain scenario should be reflected in a less frequent use, the remaining three regular scenarios scored below the average satisfaction score, indicating that some regular scenarios are particularly unsatisfactory for the trainee translators. Notably, the scenario “providing translation references for sentences or paragraphs” ranked fourth in frequency of use but last in satisfaction, suggesting that while the trainee translators had a strong demand for machine translation in their work, current LLM-based machine translation failed to meet their requirements, an issue that warrants further investigation.

In addition, the comparison between the average prompts and the prompting times of each scenario demonstrated that of the seven scenarios, five scenarios—shaded in Table 3, Column Prompting Times—including “summarizing the content of the source text”, “highlighting key terms or phrases that require special attention”, “providing background knowledge or external resources for understanding the source text”, “answering specific questions about terminology, grammar, or cultural references”, and “providing translation references for sentences or paragraphs” ranked below the overall average level. Considering their satisfaction scores, the less prompting times in scenarios including “highlighting key terms or phrases that require special attention”, “providing background knowledge or external resources for understanding the source text”, and “answering specific questions about terminology, grammar, or cultural references” suggest a high efficiency of LLMs’ performance, explaining why trainee translators have demonstrated strong satisfaction with the three scenarios. However, “summarizing the content of the source text” and “providing translation references for sentences or paragraphs”

scenarios, though requiring fewer prompts compared to the average, scored lower than the average satisfaction level. This suggests that it is possible students may have obtained outputs from LLMs that were far from satisfactory ones and thus gave up prompting after first several rounds of interactions. As for the rest two scenarios ranking above the average prompts, trainee translators' satisfaction with "identifying (and correcting) grammatical, semantic, or stylistic issues in the target text" was higher than the average level, indicating that the students had a great need for explanations on the above issues, and that this scenario was of great importance to their translation practice. In contrast, "suggesting appropriate tone, style, or register for the translation" scored below the average satisfaction level, implying that students failed to obtain satisfactory answers after multiple turns of prompts. These findings demonstrate the need to develop prompts tailored to specific tasks, with the aim of maximizing the effectiveness of the initial response.

5.3 Three scenarios where trainee translators rarely use LLMs

Meanwhile, more than 60 percent of participants reported that they had never, or rarely, asked LLMs to "examine whether the translation meets the standard of classic translation norms like 'faithfulness, expressiveness and elegance'", "provide suggestions for future translations based on past feedback", or "provide feedback from the target audience's perspective".

When asked for the reasons, over 80 percent of participants stated that they had never thought of using LLMs to "provide suggestions for future translations based on past feedback", or to "provide feedback from the target audience's perspective", both of which belong to the post-production stage. One possible explanation is that trainee translators or translation training programs do not attach great importance to this stage, despite its importance in improving the quality of a final translation product and trainee translators' competence in the long run by providing continuous feedback and suggestions to support their development. Another possibility is that trainee translators believe post-production jobs should be performed by human beings rather than LLMs and have therefore never tried to use LLMs for this stage. However, it is worth noting that the

satisfaction score for the performance of LLMs in "providing feedback from the target audience's perspective" ranked 3rd across the ten scenarios. To some extent, this suggests that LLMs are effective and useful for those who regularly use them in this scenario, proving the suitability of LLMs at the post-production stage. Raising awareness of these benefits could promote the use of LLMs in post-production among trainee translators. In addition, over 50 percent of respondents claimed that they had never thought of asking LLMs to "examine whether the translation meets the standard of classic translation norms like 'faithfulness, expressiveness and elegance'". The results indicate that trainee translators tend to pay less attention to translation norms during the production stage. One avenue for further development could be to incorporate translation norms into the design for prompt engineering, in addition to calling on translator trainers to encourage a combination of theory and practice in teaching AI-enhanced translation activities.

The satisfaction scores for two of the three scenarios—"examining whether the translation meets the standard of classic translation norms like "faithfulness, expressiveness and elegance" and "providing suggestions for future translations based on past feedback"—were relatively low. This could be attributed to poor human-AI communication and/or the current technical limitations of LLMs, as evidenced by the ranking of the prompting times, where these two scenarios were ranked among the top three. Although detailed reasons have not yet been explored, it is probable that translators may have to invest much more effort when interacting with LLMs in these situations. These results suggest the need to improve trainee translators' prompt engineering skills and fine tune LLMs to meet user expectations.

5.4 Trainee translators' commonly used LLMs

The list of LLMs used by participants, along with the frequency of their mentions in the questionnaire, is presented in Table 5.

As shown in the table, trainee translators tend to prefer open-source, general-purpose LLMs over translation-specific LLMs or LLM-integrated CAT tools. It should be noted that one student mentioned DeepL—which is not an

LLM—in the survey, indicating that there are still students preferring traditional machine translation tools rather than LLMs. One possible explanation could be students’ limited access to commercial models designed specifically for translation tasks. In addition, these commercial models are often tailored to the needs of professional translators and thus may not fulfill the expectations or

Name	Times
ChatGPT	37
Kimi	14
Cici (Doubao)	8
ERNIE Bot (Wenxinyiyan)	3
Claude	2
Deepseek	2
Tongyi Qianwen	2
Gemini	1
Grammarly	1
WPS Lingxi	1
Quark	1

Table 5: The LLMs Mentioned by Participants and the Number of Times they were Mentioned

learning needs of students.

Of the listed LLMs, ChatGPT was the most popular one. One contributing factor may have been its advanced intelligence. After GPT-4 was launched, it was tested to solve problems in various cases more effectively than the original ChatGPT and to perform tasks at a level comparable to that of human beings (Bubeck et al., 2023). Such results have boosted its reputation and popularity. Another possible reason is that, as an LLM targeting global users, ChatGPT performs better in generating English texts compared with Chinese domestic LLMs. The next two LLMs, Kimi and Cici, are both developed in China. Although they were less widely favored than ChatGPT, the frequency with which they were mentioned by participants might indicate a growing preference among trainee translators for domestic LLMs. The researcher assumes that users might be satisfied with their performance in understanding Chinese text due to the possibility that their training data is more closely aligned with Chinese culture. However, so far, no relevant studies have confirmed this assumption, nor is there any evidence on the sources of the training data used by Kimi and Cici.

5.5 Insights for the design of an LLM translator guide

The results of the questionnaire provide insights into the development of an LLM translator guide—a chatbot designed for translators’ workstations, which aggregates scenario-specific LLMs and serves as a reference tool to direct translators to the appropriate LLM for different scenarios and provides ready-to-use prompts.

The findings indicate that trainee translators rely more on LLMs during the pre-production and production stages and engage with these tools less frequently in the post-production phase. Nevertheless, an effective LLM translator guide could cover scenarios across all three stages given the potential of LLMs to assist trainee translators throughout their workflow. Therefore, a practical starting point from which to develop the initial version of the LLM translator guide would be to focus on all scenarios identified and analyzed in this study, with particular attention to the post-production stage.

The results also shed light on the design of scenario-specific LLMs. On the one hand, for scenarios where current LLMs have already met basic requirements, scenario-specific LLMs could build on popular tools such as ChatGPT and Kimi, focusing on fine-tuning the models as well as improving the design of user interfaces and, ultimately, user experience. On the other hand, for scenarios where current LLMs have so far failed to meet the needs of trainee translators, the challenge is not only to design scenario-specific LLMs but also to ensure that the LLM translator guide optimizes existing functions by incorporating guidance and examples for prompt engineering. For instance, in scenarios where trainee translators currently experience more rounds of interaction compared with less effort-cost scenarios, priority should be given to improving the prompt engineering skills of trainee translators and the ability of the LLM translator guide to interpret prompts and provide more targeted responses, which would enhance the efficiency and effectiveness of human-AI communication.

Furthermore, when designing the LLM translator guide, clear instructions explaining its functions should be included to prevent its underuse due to insufficient awareness of specific features, as seen in certain post-production scenarios. These instructions need to be

accompanied by training to ensure students understand how to use the translator guide across the pre-production, production, and post-production stages. Such preparation would not only enhance the potential of the LLM translator guide as an innovative teaching and learning tool but also better prepare trainee translators to enter the AI-integrated translation workflow and industry in the future.

5.6 Limitations and future research plan

This study has identified regular scenarios where trainee translators commonly use LLMs or rarely resort to LLMs and has surveyed their regularly utilized LLMs when carrying out their translation tasks. However, several unresolved issues remain. The first issue is exploring why LLMs including ChatGPT, Kimi, and Cici were the most popular choices among the trainee translators. It is also worth exploring why the trainee translators tended not to use LLMs at the post-production stage—whether it was because of their lack of consideration for this stage or limitations in the ability of LLMs to perform tasks effectively. There has also been insufficient investigation into trainee translators' evaluations of the performance of LLMs. While they assessed the overall performance of LLMs, the satisfactoriness of their specific functions or design in certain scenarios, as well as the disadvantages that need improvement, are still unknown. A more comprehensive understanding is therefore needed to facilitate the design of scenario-specific LLMs in the future.

Therefore, in the follow-up research, we intend to conduct focus group interviews to explore translators' use of LLMs, their evaluations, and suggestions for LLM improvement. In terms of their use of LLMs, they will first be asked to share the reasons why they prefer to use certain LLMs during their translation practice. Their answers will help address the first issue mentioned above. In this part, they will also be asked to describe how they use LLMs across the three stages, which will inform the researchers of the details concerning their interactions with LLMs and help explain why they rarely use them during the post-production stage. As for their evaluation of LLMs and suggestions for improvement, the researchers will invite participants to systematically evaluate their performance, identify the deficiencies of current LLMs, and share their opinions on how

these tools can be improved to meet their needs. Participants' answers to these two aspects will further clarify their rating of LLM performance.

Furthermore, since this study has identified the regular roles played by LLMs and the scenarios where trainee translators might use them, and given that the LLM translator guide is intended not only to serve as a specialized tool for trainee translators but also to support their translation learning, the researchers also aim to design a framework in the future to evaluate the performance of LLMs specifically for translation education from a user perspective. Drawing on existing evaluation frameworks based on user experience, the researchers are currently developing a customized framework to evaluate the use of LLMs in translation classrooms, which, broadly speaking, will be conducive to providing an AI-driven, immersive learning experience for trainee translators as well as promoting the integration of AI into translation pedagogy.

6 Conclusion

As a preliminary step towards building an LLM translator guide for trainee translators, this study has investigated the scenarios in which trainee translators rely on LLMs during their translation workflow, based on questionnaire research. The findings revealed that interactions between trainee translators and LLMs occurred mainly in the pre-production and production stages, where LLMs were used for tasks such as question answering, correction, content generation, and summarization. In contrast, the post-production stage saw less engagement with LLMs. Moreover, despite the fact that trainee translators have already started to integrate LLMs into their translation workflow, their evaluation of the performance of LLMs revealed areas for improvement.

This study has laid the groundwork for further research and development to optimize human-AI collaboration in translation. It is hoped that the final product—the LLM translator guide—will enhance the competence of trainee translators and better prepare them for human-AI collaboration in future practice. If it works, the project will be extended to design a guide for professional translators as well, to improve their work efficiency and enable them to thrive in an AI-integrated translation industry.

References

- Elisa Alonso, & Lucas Nunes Vieira. 2017. The Translator's Amanuensis 2020. *JoSTrans: The Journal of Specialised Translation*(28):345-361. http://www.jostrans.org/issue28/art_alonso.php
- Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, & Pascale Fung. 2023. A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity.arXiv:2302.04023. Retrieved February 01, 2023, from <https://ui.adsabs.harvard.edu/abs/2023arXiv230204023B>.
- Rachel Bawden, & François Yvon. Year. Investigating the Translation Performance of a Large Multilingual Language Model: the Case of BLOOM.In *Proceeding Investigating the Translation Performance of a Large Multilingual Language Model: the Case of BLOOM*, European Association for Machine Translation, pages 157-170. <https://aclanthology.org/2023.eamt-1.16/>
- Santosh Bhupathi. 2025. Role of Databases in GenAI Applications. *arXiv preprint arXiv:2503.04847*.
- Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuezhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, & Yi Zhang. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4.arXiv:2303.12712. Retrieved March 01, 2023, from <https://ui.adsabs.harvard.edu/abs/2023arXiv230312712B>.
- Sourojit Ghosh, & Srishti Chatterjee. 2024. Machine Translation, Large Language Models, and Generative AI in the University Classroom:Toward a Pedagogy of Care. In E. Monzó-Nebot & V. Tasa-Fuster (ed.), *The Social Impact of Automating Translation*, pages. Routledge
- Amr Hendy, Mohamed Abdelrehim, Amr Sharaf, Vikas Raunak, Mohamed Gabr, Hitokazu Matsushita, Young Jin Kim, Mohamed Afify, & Hany Hassan Awadalla. 2023. How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation.arXiv:2302.09210. Retrieved February 01, 2023, from <https://ui.adsabs.harvard.edu/abs/2023arXiv230209210H>.
- Martin Kay. 1980. The Proper Place of Men and Machines in Language Translation. <https://aclanthology.org/www.mt-archive.info/70/Kay-1980.pdf>
- Jiří Levý. 1967. Translation as A Decision Process. In (ed.), *To honor Roman Jakobson : essays on the occasion of his 70. birthday, 11. October 1966*, pages 1171-1182. De Gruyter Mouton.<https://doi.org/doi:10.1515/9783111349121-031>.
- Robert Martinez Carrasco, Anabel Borja Albi, & Łucja Biel. 2024. Legal translation in the face of (de)globalisation. The impact of human development, polycrises and technological disruptions in language service provision. *MonTI. Monographs in Translation and Interpreting*(16). <https://www.e-revistas.uji.es/index.php/monti/article/view/8116>
- Alan Melby. 1992. The translator workstation. In J. Newton (ed.), *Computers in Translation*, pages 147-165. Routledge.<https://doi.org/10.4324/9780203128978>.
- Shereen A Mohamed, Ashraf A Elsayed, YF Hassan, & Mohamed A Abdou. 2021. Neural machine translation: past, present, and future. *Neural Computing and Applications*, 33:15919-15931.
- Joss Moorkens, & Ana Guerbero Arenas. 2024. Artificial intelligence, automation and the language industry. In M. Gary, E.-D. Maureen, & A. Erik (ed.), *Handbook of the Language Industry*, pages 71-98. De Gruyter Mouton.<https://doi.org/doi:10.1515/9783110716047-005>.
- Joss Moorkens, Sheila Castilho, Federico Gaspari, Antonio Toral, & Maja Popović. 2024. Proposal for a Triple Bottom Line for Translation Automation and Sustainability: An Editorial Position Paper. *The Journal of Specialised Translation*(41):2-25. <https://doi.org/10.26034/cm.jostrans.2024.4706>.
- Yasmin Moslem, Rejwanul Haque, John D. Kelleher, & Andy Way. Year. Adaptive Machine Translation with Large Language Models.In *Proceeding Adaptive Machine Translation with Large Language Models*, European Association for Machine Translation, pages 227-237. <https://aclanthology.org/2023.eamt-1.22/>
- Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, & Ajmal Mian. 2023. A Comprehensive Overview of Large Language Models.arXiv:2307.06435. Retrieved July 01, 2023, from <https://ui.adsabs.harvard.edu/abs/2023arXiv230706435N>.

- Sharon O'Brien. 2024. Human-Centered augmented translation: against antagonistic dualisms. *Perspectives*, 32(3):391-406. <https://doi.org/10.1080/0907676X.2023.2247423>.
- Yuhong Peng, Huihui Huang, & Defeng Li. 2024. *New Advances in Translation Technology: Applications and Pedagogy*. Springer. <https://doi.org/10.1007/978-981-97-2958-6>.
- Nooshin Pourkamali, & Shler Ebrahim Sharifi. 2024. Machine Translation with Large Language Models: Prompt Engineering for Persian, English, and Russian Directions.arXiv:2401.08429. Retrieved January 01, 2024, from <https://ui.adsabs.harvard.edu/abs/2024arXiv240108429P>.
- Anthony Pym, & Hao Yu. 2024. *How to Augment Language Skills*. Routledge. <https://doi.org/10.4324/9781032648033>.
- Yousef Sahari, Abdu M Talib Al-Kadi, & Jamal Kaid Mohammed Ali. 2023. A cross sectional study of ChatGPT in translation: Magnitude of use, attitudes, and uncertainties. *Journal of Psycholinguistic Research*, 52(6):2937-2954.
- Mohammed Q. Shormani. 2024. Artificial intelligence contribution to translation industry: looking back and forward.arXiv:2411.19855. Retrieved November 01, 2024, from <https://ui.adsabs.harvard.edu/abs/2024arXiv241119855S>.
- Sai Cheong Siu. 2023. ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation.
- International Organization for Standardization. 2015. Translation services — Requirements for translation services of the Standard. ISO 17100:2015, International Organization for Standardization.
- International Organization for Standardization. 2024. Translation projects — General guidance of the Standard. ISO 11669:2024, International Organization for Standardization.
- Hari Venkatesan. 2023. Technology preparedness and translator training. *Babel*, 69(5):666-703. <https://doi.org/https://doi.org/10.1075/babel.00335.ven>.
- Lulu Wang, Simin Xu, & Kanglong Liu. 2024. Understanding Students' Acceptance of ChatGPT as a Translation Tool: A UTAUT Model Analysis.arXiv:2406.06254. Retrieved June 01, 2024, from <https://ui.adsabs.harvard.edu/abs/2024arXiv240606254W>.
- Yun Wang, & Zheng Zhang. 2024. The Pitfall and Relief of ChatGPT Artificial Intelligence Translation. *Chinese Translators Journal*, 45(2):95-102.
- Wordscope. *Translate your documents faster and better thanks to Artificial Intelligence*. <https://pro.wordscope.com/>
- Masaru Yamada. Year. Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability.In *Proceeding Optimizing Machine Translation through Prompt Engineering: An Investigation into ChatGPT's Customizability*, Asia-Pacific Association for Machine Translation, pages 195-204. <https://aclanthology.org/2023.mtsummit-users.19/>
- Hao Yu, & Yunyun Guo. 2024. Risk and Transcendence: an ethical analysis of ChatGPT enabling translation. *Chinese Translators Journal*, 45(4):115-122.
- Biao Zhang, Barry Haddow, & Alexandra Birch. 2023. Prompting Large Language Model for Machine Translation: A Case Study.arXiv:2301.07069. Retrieved January 01, 2023, from <https://ui.adsabs.harvard.edu/abs/2023arXiv230107069Z>.
- Wenkang Zhang, Albert W Li, & Chenze Wu. 2025. University students' perceptions of using generative AI in translation practices. *Instructional Science*:1-23. <https://doi.org/https://doi.org/10.1007/s11251-025-09705-y>.

Appendix A Questionnaire on Trainee Translators' Use of LLMs Throughout the Translation Process

- I ask LLMs to summarize the content of the source text.
 - Never or Rarely
 - Sometimes
 - Often
 - Always
 (If the participant chooses “never or rarely”, he/she will be asked to answer the following question)
 Following question: Please choose your reason (Multiple-select question).
 - I have never thought of using LLMs in this way.
 - I think LLMs' answers are not reliable.
 - I think LLMs' answers are not useful.
 - Other (Blank)
 (If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)
 Following question 1: On average, how many prompts (e.g., instructions, clarifications, or

follow-up requests) do you use when asking LLMs to summarize the content of the source text?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in summarizing content?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

2. I ask LLMs to highlight key terms or phrases that require special attention.

- ☐ Never or Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- ☐ I have never thought of using LLMs in this way.
- ☐ I think LLMs’ answers are not reliable.
- ☐ I think LLMs’ answers are not useful.
- ☐ Other (Blank)

(If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to highlight key terms or phrases that require special attention?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in highlighting key terms or phrases that require special attention?

- ☐ 1
- ☐ 2
- ☐ 3

- ☐ 4
- ☐ 5

3. I ask LLMs to provide background knowledge or external resources enabling me to understand the source text.

- ☐ Never or Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- ☐ I have never thought of using LLMs in this way.
- ☐ I think LLMs’ answers are not reliable.
- ☐ I think LLMs’ answers are not useful.
- ☐ Other (Blank)

(If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to provide background knowledge or external resources for you to understand the source text?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in providing background knowledge or external resources enabling you to understand the source text?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

4. I ask LLMs to answer specific questions about terminology, grammar, or cultural references.

- ☐ Never or Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- ☐ I have never thought of using LLMs in this way.
- ☐ I think LLMs' answers are not reliable.
- ☐ I think LLMs' answers are not useful.
- ☐ Other (Blank)

(If the participant chooses "sometimes" to "always", he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to answer specific questions about terminology, grammar, or cultural references?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in answering specific questions about terminology, grammar, or cultural references?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

5. I ask LLMs to suggest appropriate style, tone, and register of the translation.

- ☐ Never or Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(If the participant chooses "never or rarely", he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- ☐ I have never thought of using LLMs in this way.
- ☐ I think LLMs' answers are not reliable.
- ☐ I think LLMs' answers are not useful.
- ☐ Other (Blank)

(If the participant chooses "sometimes" to "always", he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking

LLMs to suggest appropriate style, tone, and register of the translation?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in suggesting appropriate style, tone, and register of the translation?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

6. I ask LLMs to provide translation references for sentences or paragraphs.

- ☐ Never or Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(If the participant chooses "never or rarely", he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- ☐ I have never thought of using LLMs in this way.
- ☐ I think LLMs' answers are not reliable.
- ☐ I think LLMs' answers are not useful.
- ☐ Other (Blank)

(If the participant chooses "sometimes" to "always", he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to provide translation references for sentences or paragraphs?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in providing translation references for sentences or paragraphs?

- ☐ 1
- ☐ 2
- ☐ 3

- 4
- 5

7. I ask LLMs to identify (and correct) grammatical, semantic, or stylistic issues in the target text.

- Never or Rarely
- Sometimes
- Often
- Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- I have never thought of using LLMs in this way.
- I think LLMs’ answers are not reliable.
- I think LLMs’ answers are not useful.
- Other (Blank)

(If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to identify and correct grammatical, semantic, or stylistic issues in the target text?

- 1-5
- 6-10
- 11-15
- 16-20
- Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in identifying and correcting grammatical, semantic, or stylistic issues in the target text?

- 1
- 2
- 3
- 4
- 5

8. I ask LLMs to examine whether the translation meets the standard of classic translation norms like “faithfulness, expressiveness and elegance”.

- Never or Rarely
- Sometimes
- Often
- Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- I have never thought of using LLMs in this way.
- I think LLMs’ answers are not reliable.
- I think LLMs’ answers are not useful.
- Other (Blank)

(If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to examine whether the translation meets the standard of classic translation norms like “faithfulness, expressiveness and elegance”?

- 1-5
- 6-10
- 11-15
- 16-20
- Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in examining whether the translation meets the standard of classic translation norms like “faithfulness, expressiveness and elegance” ?

- 1
- 2
- 3
- 4
- 5

9. I ask LLMs to provide feedback from the target audience’s perspective.

- Never or Rarely
- Sometimes
- Often
- Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- I have never thought of using LLMs in this way.
- I think LLMs’ answers are not reliable.
- I think LLMs’ answers are not useful.
- Other (Blank)

(If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to provide feedback from the target audience’s perspective?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in providing feedback from the target audience’s perspective?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

10. I ask LLMs to provide suggestions for my future translations.

- ☐ Never or Rarely
- ☐ Sometimes
- ☐ Often
- ☐ Always

(If the participant chooses “never or rarely”, he/she will be asked to answer the following question)

Following question: Please choose your reason (Multiple-select question).

- ☐ I have never thought of using LLMs in this way.
- ☐ I think LLMs’ answers are not reliable.
- ☐ I think LLMs’ answers are not useful.
- ☐ Other (Blank)

(If the participant chooses “sometimes” to “always”, he/she will be asked to answer the following two questions)

Following question 1: On average, how many prompts (e.g., instructions, clarifications, or follow-up requests) do you use when asking LLMs to provide suggestions for your future translations?

- ☐ 1-5
- ☐ 6-10
- ☐ 11-15
- ☐ 16-20
- ☐ Over 20

Following question 2: On a scale of 1 to 5, with 5 being the highest score, how would you rate the performance of LLMs in providing suggestions for your future translations?

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

11. If there are other scenarios that are not mentioned above, please write them down.

(Blank)

12. Please write down the names of your most commonly used LLMs.

(Blank)

Appendix B Weighted Average Scores for Three Metrics

Scenarios	Weighted Average for Three Metrics		
	Frequency Score	Prompting Times	Satisfaction Score
Summarizing the content of the source text	1.10	1.36	3.25
Highlighting key terms or phrases that require special attention	0.76	1.32	3.41
Providing background knowledge or external resources for understanding the source text	1.51	1.53	3.64
Answering specific questions about terminology, grammar, or cultural references	1.71	1.46	3.77
Suggesting appropriate style, tone, and register for the translation	0.80	1.78	3.35
Providing translation references for sentences or paragraphs	1.22	1.53	3.13
Identifying (and correcting) grammatical, semantic, or stylistic issues in the target text	1.27	1.69	3.47
Examining whether the translation meets the standard of classic translation norms like “faithfulness, expressiveness and elegance”	0.56	1.93	3.14
Providing feedback from the target audience’s perspective	0.39	1.67	3.58
Providing suggestions for future translations based on past feedback	0.51	1.73	3.33