

LLM-Empowered Translation Teaching: A Case Study of the Chinese-English Translation Course

MA Jing

Qingdao University of Science and Technology, Qingdao, China

This study proposes a three-stage teaching model—pre-translation preparation, human-machine collaboration, and post-translation reflection—in the context of the Chinese-English translation course, as a way to integrate large language models into translation teaching. The findings indicate that this approach enhances students’ awareness of terminological consistency, refines their sensitivity to cultural appropriateness, and strengthens their critical engagement with artificial intelligence (AI)-generated translations. However, the model shows clear limitations when the source text carries a heavy cultural load. The paper also discusses the practical challenges observed during implementation and suggests corresponding coping strategies, with the aim of offering useful references for pedagogical reform in similar courses.

Keywords: large language models, human-machine collaboration, action research, Chinese-English translation

Introduction

The rapid advance of large language models (LLMs), such as ChatGPT and DeepSeek, is reshaping both the translation industry and translator training. With the field moving from machine translation post-editing toward artificial intelligence (AI)-assisted post-editing, educators are being pushed to rethink their priorities (Wang & Liu, 2025). As Wang Kefei (2024) noted, AI handles routine translation tasks well enough, but work that draws on human judgment, cultural sensitivity, or creative thinking still requires a human translator. This suggests that the traditional classroom model, in which students translate independently and receive instructor feedback, may no longer be adequate. That model served its purpose when students worked as the sole translators. But now that AI can produce translations of reasonably high quality in seconds, the challenge for translation teachers is no longer whether to use these tools, but how to help students engage with them thoughtfully—knowing when to trust the output, when to question it, and when to set it aside. These questions are becoming increasingly unavoidable for educators and researchers alike.

Considerable efforts have been made to explore the intersection of LLMs and translation teaching. Some studies have examined the performance of AI translation tools, others have looked into their application in developing teaching resources, and still others have addressed ethical concerns and the risks of student over-reliance (Sheng, 2023). Empirical work has also begun to emerge. Zhang and Peng (2025), for instance, found

Acknowledgements: This paper is granted by the projects “LLM-Supported Translation Teaching Reform in Higher Education: An Exploratory Study (2024),” “Fostering Intelligent Competence in Translation Teaching: A Language Intelligence Perspective (2023),” and “Foreign Language Teachers’ EdTech Competence and Platform Development in the Age of AI” (2023).

MA Jing, associate professor, School of Foreign Languages, Qingdao University of Science and Technology, Qingdao, China.

that while LLMs handle structured English-to-Chinese tasks fairly well, their consistency drops noticeably on Chinese-to-English tasks involving cultural or semantic subtleties. Wang and Zhang (2025) further noted that despite the benefits of LLMs in providing resources and instant feedback, these tools also bring risks that require careful handling in the classroom. Yet, for all this activity, systematic exploration of teaching models remains scarce, with most studies staying theoretical and lacking empirical grounding in actual practice.

Based on these considerations, this study takes the Chinese-English translation course for English majors at a university as its practical setting, adopts an action research methodology, and attempts to address three core questions:

1. How can LLMs be effectively embedded into the translation teaching workflow?
2. What characteristics do students exhibit during this human-machine collaborative training process?
3. What difficulties arise in practice and how can they be effectively addressed?

Design and Implementation of the Practical Scheme

Design Principles

The design of this scheme rests on three fundamental premises.

First, LLMs, despite their impressive capabilities, do not possess the cultural judgment or aesthetic sensibility of human translators and cannot replace the translator's subjective decision-making;

Second, the primary objective of translation teaching is not to instruct students in the use of any particular tool, but to cultivate their translation judgment and lifelong learning capabilities;

Third, the introduction of any teaching tool must be accompanied by appropriate pedagogical norms and evaluation mechanisms.

Building on these premises, the scheme establishes three implementation principles.

The first is "human first, machine second." Students must complete their own initial translation before consulting any AI-generated output, ensuring that the independent thinking phase is not cut out;

The second is "human-controlled AI," meaning AI serves as a reference and auxiliary tool, while the final translation decisions rest with the students themselves;

The third principle, "process-oriented," holds that teachers should consider not only the quality of the final translation product, but also the quality of judgment students demonstrate in comparing, analyzing, and revising during the collaborative process.

Participants and Implementation Conditions

The practice was conducted with 32 third-year English majors specializing in translation at a university. These students had already completed foundational English courses and introductory translation courses, and thus possessed basic bilingual transfer competence sufficient for the tasks involved. Each student had access to a computer or tablet device with a stable internet connection, and had completed pre-course training on LLMs, including ethical considerations, data privacy, and appropriate citation.

Teaching Procedure

The teaching procedure comprises three phases: pre-translation preparation, classroom discussion, and post-class reflection.

During the pre-translation preparation phase, students complete their initial translation drafts independently without using AI tools. In this phase, students are required to highlight difficulties encountered during translation,

including terms they are unsure of, difficult sentence structures, and culturally loaded expressions. Upon completion, students submit their drafts to the platform Xuexitong. The instructor reviews the submitted translations, identifies common problems, and adjusts the focus of classroom discussion accordingly.

The classroom discussion phase consists of five steps.

First, the instructor organizes small-group discussions on difficulties identified in the initial translation, allowing peer assistance to resolve some issues.

In the second step, the teacher demonstrates how to craft effective prompts for different translation task types, highlighting key parameters, such as “task type,” “target audience,” “style requirements,” and “terminological constraints.” This echoes Sheng’s (2023) observation that only precise and well-structured prompting enables LLMs to produce genuinely useful translations.

In the third step, student groups use their self-designed prompts to obtain reference translations via LLMs.

In the fourth and most crucial step, each group conducts a three-way comparison among their own translation, the AI-generated version, and their initial draft, analyzing differences in terms of terminological consistency, stylistic appropriateness, and cultural information transfer, and discussing the rationale for their choices.

Finally, each group then reports their findings to the class after which the instructor provides a synthesis and supplementary comment, with particular attention to typical issues in AI-generated translations—such as simplification of cultural information and stylistic homogenization.

During the post-class reflection phase, students revise their translations based on classroom discussion outcomes and compose brief reflective logs addressing three questions:

1. What assistance AI provided for this translation task?
2. Which elements of the AI-generated translation were judged inappropriate and why?
3. How they plan to refine their prompting strategies or translation approaches in similar future tasks?

The instructor reviews and provides written feedback on both the revised drafts and the reflective logs.

Representative Teaching Cases

A case involving culturally loaded expressions is particularly illustrative. The source text contained Chinese proverbs, such as “留得青山在，不怕没柴烧” (While green hills remain, one need not worry about firewood) and “树大招风” (A tall tree catches the wind). During pre-translation preparation, most students chose literal translation, rendering “树大招风” as “A tall tree catches the wind.” In the classroom session, students obtained multiple reference versions via LLMs, including literal, free, and annotated translations with explanatory notes. Group discussions centered on several key questions:

1. Whether literal translation would effectively convey the implied meaning to English readers?
2. Whether free translation would sacrifice cultural distinctiveness in the process?
3. Whether there exists a more effective strategy?

After comparison, most groups preferred a “literal translation plus brief explanation” approach, which preserved the cultural imagery while simultaneously bridging the comprehension gap for target readers. In their reflective logs, students commonly noted that the most valuable aspect of such translation tasks was not obtaining a “correct answer” from AI, but being compelled, in the process of comparing multiple versions, to think through the rationale behind “Why this translation is better?”

In terminological standardization training, a public policy text containing numerous specialized terms was selected. In the initial translation phase, students handled terminology rather casually—the same concept was

rendered with different equivalents at different points in the text. By consulting AI-generated translations, students observed that AI exhibited a higher degree of terminological consistency. Following classroom discussion, students progressively recognized the importance of terminology management and, in subsequent translation assignments, consciously constructed term glossaries.

Results and Reflections

Observed Improvements

After one semester of sustained practice, students showed notable improvements across several dimensions.

First, their translation decision-making became more rationalized. Early in the semester, students relied heavily on intuition in their translation work, and when faced with AI-generated versions, they tended to choose whichever “read more smoothly.” By the middle and later stages of the semester, students began to invoke concepts, such as “terminological consistency,” “register matching,” and “cultural adaptability” to justify their judgments, marking a shift from experiential intuition to reasoned analysis.

Second, their attitudes toward AI tools became more critical. An early-semester survey revealed that approximately 60% of students expressed “high trust” in LLMs, believing their translation quality approached or even exceeded human performance. By the end of the semester, however, most students could identify typical AI translation problems, including stiff and unnatural handling of rhetorical language, frequent omission or oversimplification of cultural information, and a tendency toward stylistic homogenization. They had also developed the understanding that AI output should not be blindly trusted but rather critically evaluated in each instance.

Third, their awareness of terminological and cultural issues strengthened significantly. Students’ reflective logs made frequent reference to terminology consistency and cultural appropriateness, suggesting that these two dimensions had become internalized.

Major Challenges

Several challenges were encountered during implementation.

The first was student over-reliance. Despite the explicit “human first” requirement spelled out from the start, a handful of students still turned to AI translations during out-of-class preparation without doing their own thinking first. Rules alone cannot fully solve this, but tighter process management can help. For example, instructors can require students to submit drafts with visible revision records, call on them randomly in class to explain their choices, and give weight to the thinking process in their final grade.

Classroom time constraints posed another substantial challenge. Human-machine translation comparison requires sufficient time, yet the 100-minute class session often proves too short to cover lectures, discussions, presentations, and practice all at once. One strategy for addressing this constraint involves delivering certain content via micro-lectures. Through pre-class video instruction, students can acquire operational skills such as prompt design. This approach, in turn, frees up classroom time for more interactive discussion and hands-on training.

Assessment criteria have yet to catch up with pedagogical changes. Traditional grading still prioritizes accuracy, fluency, and completeness of the final translation version. A key challenge is how to fairly evaluate the judgment and strategic choices students make during human-machine collaboration. The current solution is to weight process-oriented assessment more heavily, factoring in classroom participation, reflective journals, and comparative analysis reports. But this approach is still not perfect and needs more refinement.

Reflections on LLM-Assisted Translation Teaching

Several insights from this semester-long practice merit sustained attention.

LLMs function more as a reference than as an authoritative model. In traditional translation teaching, the teacher's reference translation often served implicitly as the standard that students were expected to aim for. LLMs now can rapidly generate multiple versions for students to compare against one another, against their own translations, and against the teacher's version. When faced with several acceptable translations, students must exercise their own judgment and make informed choices. It is precisely this act of choosing that lies at the core of developing translation decision-making competence. In this sense, LLMs are shifting translation teaching away from delivering "standard answers" and toward cultivating judgment.

The demands on teachers are paradoxically higher, not lower. The quality of classroom discussion after introducing LLMs depends significantly on how the teacher designs analytical frameworks and guides students toward the deeper issues beneath surface-level AI output. Teachers need to know not just what AI does well, but also where it tends to fail. More importantly, they have to turn those technical weaknesses into discussion questions that genuinely engage students in meaningful reflection.

The objectives of translation teaching may need fundamental rethinking. If LLMs can already handle a significant portion of translation work competently, it is worth asking whether teaching the same content in the same way still makes sense. Tasks that AI performs well, such as terminology retrieval, grammatical correction, and basic structural rendering, can be increasingly shifted to students' independent work. Teachers are better off investing their time in areas where AI clearly falls short, such as in-depth cultural interpretation, discourse-level stylistic judgment, and discussions of translation ethics. In other words, translation teaching may be gradually moving away from skill-based training and toward judgment-based cultivation.

Conclusion

Based on a semester-long action research in the Chinese-English translation course, this study examines the integration of LLMs into translation teaching. The findings indicate that LLMs can strengthen students' translation judgment, terminological awareness, and cultural sensitivity, provided that sound pedagogical guidelines and evaluation criteria are established. At the same time, challenges, such as student over-reliance, classroom time constraints, and lagging assessment criteria require ongoing attention.

Given the small sample size and the absence of systematic quantitative data, the findings are best viewed as exploratory rather than generalizable. Larger-scale studies with pre-test and post-test comparisons would help validate these findings. Moreover, as LLM technology continues to advance rapidly, this is a question that will demand ongoing attention.

References

- Lv, Q. X., & Jiang, Z. K. (2025). The impact of GenAI-Instructor collaborative feedback on learner engagement and revision quality: An empirical study based on literary translation post-editing. *Contemporary Foreign Language Studies*, 25(5), 156-169.
- Sheng, H. (2023). Analysis of the application prospects of Large Language Model ChatGPT in translation teaching. *Language and Education Studies*, 7(4), 13-18.
- Wang, H. L., & Zhang, Z. Z. (2025). Transformations and ethical challenges of translation teaching driven by large language models. *Language and Education Studies*, 9(2), 15-21.
- Wang, H. S., & Liu, S. J. (2025). From MTPE to AIPE: Evolution of translation modes in the GenAI era and its implications for translation education. *Shandong Foreign Language Teaching*, 46(3), 111-121.

- Wang, H. S., & Xie, F. (2024). A study on the innovation of practical models in translation education driven by large language model technology. *Chinese Translators Journal*, 45(2), 70-78.
- Wang, K. F. (2024). What translation can and cannot do in the intelligent era. *Journal of Foreign Languages*, 45(1), 5-9.
- Wang, Z., & Zhang, Z. (2024). Hidden concerns and solutions of ChatGPT artificial intelligence translation. *Chinese Translators Journal*, 45(2), 95-102.
- Zhang, J., & Peng, S. R. (2025). An empirical study on AI-driven evaluation of student translations using large language models. *Contemporary Foreign Language Studies*, 5, 85-96.
- Zhou, Z. L. (2023). The application of ChatGPT in translation teaching: Changes, challenges, and Countermeasures. *Journal of Beijing International Studies University*, 45(5), 134-146.