GenAI in Written Translation Education: A Systematic Review of Challenges and Pedagogical Pathways

Ying Wu[12]

[1]*Graduate School of Translation & Interpretation, Beijing Foreign Studies University, Beijing, China*

[2]*School of Foreign Languages, Wenhua College, Wuhan, China*

wuying_wy@whc.edu.cn

Abstract. Since late 2022, generative artificial intelligence has steadily found its way into written translation classrooms, setting off a wave of discussion about what these tools mean for how translation is taught. This paper surveys the recent literature—spanning both English and Chinese academic journals over roughly four years—to map out the specific challenges GenAI has created for written translation education and the teaching strategies scholars have proposed in response. The review finds that two concerns dominate the conversation. One is the breakdown of feedback loops: AI-generated feedback tends to stay at the surface of language and misses the deeper, context-sensitive judgment that written translation instruction depends on. The other is the erosion of foundational skills—students leaning too heavily on large language models before doing their own analytical work. Beyond these, several additional themes run through the literature: the difficulty novice translators face in catching subtle errors embedded in fluent-looking AI drafts, the anxiety that pervasive AI use generates around professional identity and ethical boundaries, the tendency of existing AI scoring tools to favor safe, formulaic output while penalizing creative or stylistically distinctive translation, and the well-documented but under-researched problem of cultural transfer failures in LLM-generated texts. On the response side, curriculum and pedagogical redesign is the most widely endorsed direction, with growing interest in building translation-specific AI literacy and in designing structured human-AI collaborative workflows. The review also finds that scholars writing in Chinese and in English emphasize different priorities—particularly around whether translator role reshaping should be institution-led or learner-driven—and that the teaching of cultural transfer competence in an AI-mediated environment remains at an early stage in both research communities.

Keywords: GenAI, written translation education, systematic review, pedagogical pathways, human-AI collaboration

1. Introduction

Generative AI tools have spread through education faster than institutions could plan for, and written translation programmes have felt the shake-up more than most. Translation teaching depends on

something specific: an instructor reads a student's draft, identifies not just what is wrong but why, and communicates that reasoning in terms the learner can absorb. When students walk into class already carrying tools that produce fluent drafts with a single prompt, that feedback relationship starts to fray. The shift from Machine Translation Post-Editing to AI Post-Editing has happened in roughly three years, and it is not simply a tool upgrade. It forces a rethink of what competencies need to be taught, how ethical choices are made in the translation process, and what a classroom workflow should even look like. Students now use large language models that generate polished text at speed, though the output can be subtly wrong or culturally off-key in ways that only an experienced reader catches. Instructors, meanwhile, are left deciding what still deserves to be taught and how to assess work when the boundary between student effort and machine output has blurred. Wang and Xie [1] put it plainly: bringing LLMs into translation education is not a tweak—it asks for structural rethinking.

Research on GenAI in education has grown quickly and now covers a wide range of subjects—language learning, automated assessment, teacher training, and more. But once the lens is narrowed to written translation education, the picture thins considerably. Work that might be relevant is scattered across several fields: general English-language education, where GenAI is often treated as a writing or vocabulary tool; professional translator training, which tends to focus on workflow efficiency rather than pedagogy; and the more specialised literature on translation teaching itself, where only some studies actually centre GenAI as a core concern rather than a footnote. When these strands are brought together, the collective insight is smaller than the individual parts suggest. The studies that remain after filtering address important questions, but they do so in isolation—one looks at ethical risks, another at curriculum design, a third at assessment—without much effort to connect findings into a broader account of what these tools mean for translation classrooms. At the same time, the English-language and Chinese-language literatures have developed along largely separate tracks, each foregrounding priorities that the other barely touches: some scholars focus on building individual AI literacy [2], while others argue for institutional-level reform [3]. A systematic review can align these scattered pieces, identify what the literature as a whole says about the challenges GenAI presents, and map the responses educators have proposed.

To do that, this review sets out to answer two straightforward questions. RQ1: What specific, translation-related challenges has GenAI introduced into written translation education? RQ2: What pedagogical pathways have researchers put forward to meet those challenges? The analysis draws on recent publications from CNKI core journals and the Web of Science and employs a structured coding framework to capture both the challenges and the responses, along with how frequently each appears and where English and Chinese scholarship see the landscape differently.

2. Methodology

The review followed a structured workflow in three steps: first, defining what kinds of studies would and would not be included; second, coding the selected articles to extract challenges and strategies; and third, running a frequency analysis to compare patterns across the English and Chinese subsets. The following sections walk through each step in turn.

2.1. Inclusion and exclusion criteria

To build a sample that was both relevant and methodologically sound, this review applied three filters. The first was substantive: a paper had to engage with GenAI tools—ChatGPT, other LLMs, or comparable systems—as a central focus, not as a passing mention. The second was disciplinary:

the educational context had to be written translation specifically, since interpreting places different cognitive and technological demands on learners. The third was pedagogical: each study needed to make a clear contribution to how translation is taught or learned. Benchmarking studies that only measured AI output quality, and research on professional translation workflows outside a teaching context, did not meet this third filter and were set aside.

What was excluded matters as much as what was kept. Studies on interpretation teaching were removed. So was the broader literature on general language education—vocabulary acquisition, grammar instruction, and the like—because the cognitive processing involved in written translation is distinct enough that findings from those domains do not transfer neatly. Professional translator training was also excluded: although closely related, that literature tends to focus on workplace upskilling and industry certification rather than the classroom-based teaching and learning processes that the research questions of this review target. Pure technical evaluations and studies limited to professional environments outside any instructional setting were set aside for the same reason.

Searches were run across two databases: the CNKI core journals and the Web of Science core collection. The time window spanned November 2022—the month ChatGPT launched and the academic conversation around GenAI took off—through April 2026. After applying all inclusion and exclusion filters, 48 articles remained, 16 published in English and 32 in Chinese. The yearly spread of these publications is shown in Figure 1.

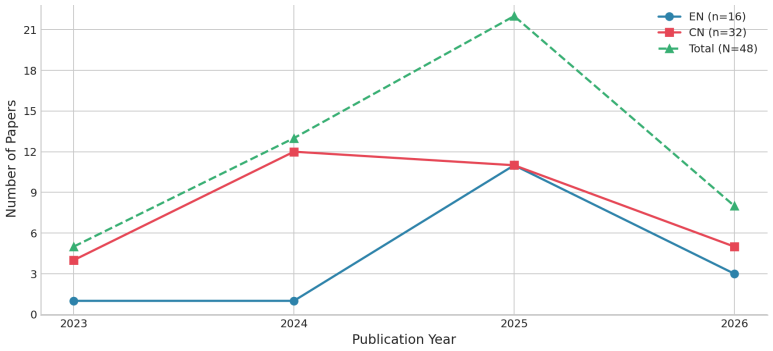

Figure 1. Annual distribution of reviewed literature (N=48)

2.2. Coding procedure

Coding proceeded through three connected phases, each building on the previous one. The initial phase was an immersive full-text reading of all 48 articles. The goal at this point was to extract every passage that described a GenAI-related challenge tied to translation education, or that laid out a teaching strategy designed in response. No attempt was made to classify or consolidate at this stage; the output was simply a large set of text segments marked by source and page number.

The consolidation phase came next. Working inductively from the extracted segments, related codes were grouped and regrouped, duplicates were merged, and a framework began to take shape around the core activities that translation teaching depends on—giving feedback, building skills, designing assessments, navigating ethics, managing post-editing workflows, and handling cultural content. The result, shown in Table 1, is a two-part taxonomy: six problem categories (P1 through P6) and six strategy categories (S1 through S6). The numbering within each list reflects how often each item appeared in the data. Because challenges and strategies were coded and counted separately, the numbers do not map across columns; P2 and S4 both concern cultural transfer competence, for instance, but cultural transfer surfaces far more often as a challenge than as a target of pedagogical proposals.

Table 1. Coding framework for challenges and strategies

Challenges (P1-P6, by descending frequency)	Pedagogical Pathways (S1-S6, by descending frequency)
P1: Limitations in Translation Quality Assessment	S1: Cultivation of Translation Literacy
P2: Deficiencies in Cultural Transfer Competence	S2: Human-AI Collaborative Translation Teaching Models
P3: Risk of Translation Competence Degradation	S3: Reconstruction of Translation Quality Assessment Systems
P4: Dilemmas in Post-Editing	S4: Cultivation of Cultural Transfer Competence
P5: Translation Ethics and Professional Identity	S5: Innovation in Curriculum and Pedagogical Models
P6: Failure of Translation Teaching Feedback Mechanisms	S6: Reshaping Translator Roles and Professional Development

The third phase was quantitative. Each of the 48 papers was cross-referenced against the taxonomy to produce a frequency count for every category. These counts were then split by language community—English (n=16) and Chinese (n=32)—so that patterns of convergence and divergence between the two research traditions could be examined directly rather than guessed at.

3. Results

The results are organized around the two research questions. Section 3.1 presents the challenges GenAI introduces into written translation education, drawing on the six problem categories identified through the coding process. Section 3.2 turns to the pedagogical strategies that researchers have proposed in response, again structured around the six strategy categories that emerged from the analysis.

3.1. Challenges introduced by GenAI

The analysis identified six distinct problem domains that GenAI introduces into written translation education. These challenges vary both in how frequently they are discussed and in how much emphasis they receive from English versus Chinese researchers as illustrated in Figure 2. The subsections below present each challenge in order of its overall prevalence in the literature. Two challenges stand out: the breakdown of feedback mechanisms and the risk of competence degradation. Cultural transfer competence, by contrast, draws far less attention than its practical importance would suggest.

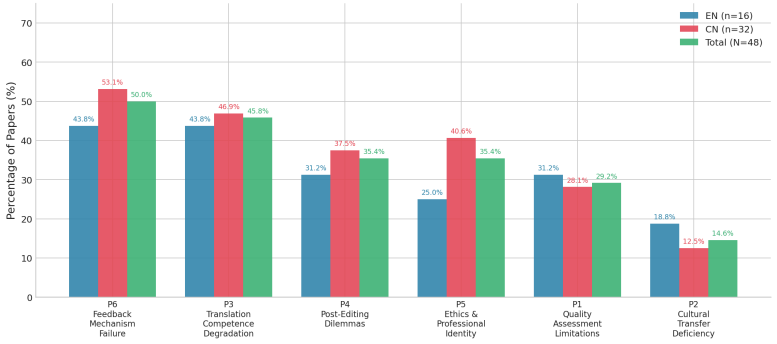

Figure 2. Frequency of translation-specific challenges (N=48)

3.1.1. Failure of translation teaching feedback mechanisms (P6)

Half of the articles in this review—24 out of 48—identify broken feedback as the most serious problem GenAI has created for translation teaching. The concern registers in 53.1% of Chinese papers (17 out of 32) and 43.8% of English ones (7 out of 16). What exactly is breaking? Translation instructors have traditionally relied on nuanced, context-rich feedback to guide students through the cognitive challenges of rendering meaning across languages. GenAI tools, by contrast, tend to stay at the surface: they catch lexical errors and awkward syntax, but they do not engage with semantic depth, cultural resonance, or the conventions of particular genres and disciplines. This mismatch is not merely annoying. When AI-generated suggestions conflict with what an instructor is trying to teach, students end up confused, and the feedback loop that drives skill development stops functioning. The problem is compounded by the fact that LLM feedback does not align with the multidimensional quality metrics (MQM) used in advanced training, which means even technically correct AI feedback can push students in directions that their coursework is designed to steer them away from. Jiao et al. [4] and Zhang et al. [5] have documented this pattern across different instructional settings, showing that the co-presence of AI and human feedback does not enrich learning when the two sources pull in opposite directions.

3.1.2. Risk of translation competence degradation (P3)

A second concern runs through nearly half the sample. Of the 48 papers, 22 (45.8%) warn that students who lean too heavily on GenAI may lose the very skills that translation training is meant to build. English and Chinese scholars flag this at almost identical rates: 43.8% of English papers and 46.9% of Chinese papers. The worry is not abstract. When a student encounters a difficult sentence and reaches for ChatGPT before working through the problem independently, the cognitive processes that underpin translation competence—analytical reading, recursive problem-solving, weighing alternative renderings—get short-circuited. Over weeks and months, the habit of deferring to AI replaces the habit of thinking through a translation problem. Wang and Xie [1] and Zhang [6] both argue that the practical consequence of this pattern is not just weaker student translations but a professional pipeline that produces editors rather than translators—people who can polish machine output but cannot generate reliable translations from their own linguistic and cultural knowledge.

3.1.3. Dilemmas in post-editing (P4)

Seventeen papers, or 35.4% of the sample, focus on problems that are specific to the post-editing process itself. Chinese scholars raise this issue somewhat more often than their English counterparts (37.5% versus 31.2%), though the asymmetry is modest. The dilemma takes a form that will be familiar to anyone who has used AI for translation: GenAI produces fluent, readable text that on first glance looks ready to use. Closer inspection, however, often reveals problems that fluency hides. Semantic shifts can be subtle—a connotation changed, a register dropped, a cultural reference flattened. Hallucinations that sound plausible because they are grammatically clean can slip past a novice reader. Several of the studies reviewed here report that students sometimes spend more effort correcting AI-generated errors than they would have spent translating the passage themselves. If that is the case, then the efficiency argument for AIPE—the claim that AI-assisted workflows save time—merits serious qualification. Geng and Hu [7] and Yao et al. [8] offer empirical data on this point, with the latter using a multi-method approach to measure the actual cognitive and temporal costs of post-editing AI output.

3.1.4. Translation ethics and professional identity (P5)

Ethics and identity concerns also surface in 35.4% of the papers (17 out of 48), but here the split between communities is sharp. Chinese scholars devote substantial attention to this area (40.6%, or 13 of 32 papers), while English-language researchers give it notably less space (25.0%, or 4 of 16). The Chinese literature tends to address several dimensions at once. One thread concerns ideological bias: LLMs trained on particular corpora may produce translations that distort politically or culturally sensitive material in ways that a human translator would recognize and correct. A second thread involves data privacy—the practical question of what happens to student translations when they are processed through commercial AI platforms that may store or train on user inputs. A third thread, more philosophical than technical, asks what it means for a translation student's sense of professional purpose when the technology can do much of what they are learning to do. Li [9] and Zhu and Wang [10] engage these questions directly, with the former mapping the shifting competency landscape and the latter probing the epistemological assumptions that underpin translation as a technology-mediated activity.

3.1.5. Limitations in translation quality assessment (P1)

Fourteen papers—29.2% of the sample, split fairly evenly between English (31.2%) and Chinese (28.1%) sources—examine the use of GenAI for assessing translation quality. The problems they identify fall into two categories. Technically, LLM-based scoring is unreliable: the same translation can receive meaningfully different scores across multiple runs, and the systems offer little diagnostic granularity about why a particular error occurred. This makes them hard to use as teaching tools, since feedback that cannot explain itself does not help a student improve. Conceptually, the issue runs deeper. LLMs apply template-driven criteria that reward safe, formulaic translations and penalize creative or stylistically distinctive choices. They perform poorly on culturally embedded texts where quality cannot be reduced to a checklist. Lu and Feng [11] and Zhang and Peng [12] provide empirical evidence for both the technical instability and the conceptual narrowness of AI-based assessment in translation education.

3.1.6. Deficiencies in cultural transfer competence (P2)

Only seven papers in the entire sample—14.6%—take up the question of what GenAI does to cultural transfer, and the numbers are thin on both sides of the language divide: 18.8% of English papers (3 out of 16) and 12.5% of Chinese papers (4 out of 32). This is, by some distance, the least-discussed challenge in the dataset, and the gap between its practical importance and its research coverage is hard to overstate. LLMs consistently mishandle culturally specific references, locally grounded idioms, and politically sensitive terminology. Translation professionals spend a significant portion of their working lives navigating exactly these challenges. Yet the educational literature has almost nothing to say about how to prepare students for this dimension of AI-assisted translation. Mu and Zheng [13] are among the few to name the problem explicitly, noting that while many papers acknowledge in passing that LLMs have cultural limitations, almost none propose systematic classroom interventions. The silence is itself a finding worth reporting.

3.2. Pedagogical pathways proposed in response

The strategies proposed in the reviewed literature fall into six groups, presented below in order of how frequently they appear, as shown in Figure 3. The split between English and Chinese priorities is also present in the response data, though—importantly—it does not always take the same shape as it did for the challenges.

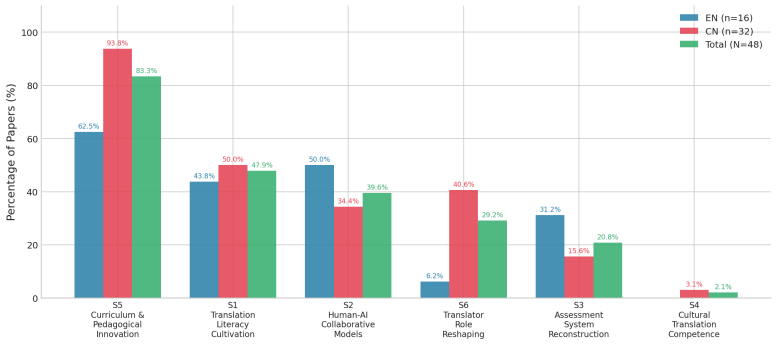

Figure 3. Frequency of pedagogical pathways (N=48)

3.2.1. Innovation in curriculum and pedagogical models (S5)

Eighty-three percent of the papers in this review—40 out of 48—call for curriculum reform. That level of consensus is rare in any academic literature, and it tells you something about how urgently translation educators feel the ground shifting. The Chinese literature is especially emphatic: 93.8% of Chinese papers (30 out of 32) make the case, compared to 62.5% of English papers (10 out of 16). What kind of reform are they asking for? The common thread is a move away from teaching models organized around the final translation product—did the student get it right?—and toward models that make the process visible: how students think, where they get stuck, what they do when AI gives them an answer that looks right but is not. Several studies propose practical steps. Move technology courses to Year 1 rather than Year 3. Design materials that can be revised semester by semester as AI tools change. Build classroom ecosystems in which students, teachers, and AI interact through structured workflows that were designed for AIPE from the start, rather than layered onto a curriculum built for a pre-AI world. Hu [3] writes from an institutional policy perspective; Wang et al [14]. focus on the classroom. Both see the same problem: the curriculum was not built for this, and it needs to be rebuilt.

3.2.2. Cultivation of translation literacy (S1)

Translation literacy—a term that has gained currency in the last few years—combines what translators have always needed (language competence) with what GenAI now demands (the ability to work critically alongside machine outputs). Twenty-three of the 48 papers, or 47.9%, recommend building this literacy explicitly into the curriculum. Unlike the sharp regional gaps seen elsewhere in the data, this one attracts balanced attention: 43.8% of English papers and 50.0% of Chinese papers. The shared argument, expressed in different ways across the two communities, is that fencing AI out of the classroom is not a sustainable strategy. Students will use these tools regardless of what instructors permit. The better bet is to equip them to use the tools well. That involves a set of skills that were not on the translation syllabus even five years ago. Students need to learn how to read AI output skeptically, recognizing when fluent text is factually wrong or culturally off-key. They need

to understand, at least in broad terms, why LLMs produce hallucinations in the first place—not as a technical deep dive, but as a conceptual orientation that prevents naive trust. They also need practical command of prompt engineering and recursive querying, the iterative back-and-forth through which a skilled user refines AI output rather than accepting the first draft. Zhang and Doherty [2] and Wang and Luo [15] have each developed frameworks for defining and assessing this literacy, with the former emphasizing empirical measurement and the latter focusing on the conceptual architecture.

3.2.3. Human-AI collaborative translation teaching models (S2)

The idea of building structured human-AI partnerships into translation teaching appears in 39.6% of the papers (19 out of 48). English-language scholars endorse the approach more often than their Chinese counterparts—50.0% versus 34.4%—though the difference is less about whether collaboration should happen than about how much institutional scaffolding it requires. The shared premise across these proposals is that GenAI functions best as a cognitive partner, not a substitute. The practical question is about timing and placement. Some researchers argue for what gets called a guided learning sequence: the student works through the source text independently first—identifying translation problems, making initial lexical and syntactic choices, flagging cultural references—and only then consults an AI tool. The point of front-loading the independent work is to protect the cognitive processes that translation training is meant to build, while still giving students experience with AI-integrated workflows. Other proposals take a different approach to the same problem, designing pipelines in which AI produces an initial draft but reserving human intervention for the stages where judgment matters most: semantic decisions, cultural calibration, stylistic choices. The empirical evidence, while still thin, offers some support for the guided learning model. Tian et al. [16] found that introducing AI after independent analysis produced measurable cognitive process benefits. Wang and Zhang [17], writing from a Chinese context, reached a complementary conclusion: the collaboration only works when the protocol is explicit and consistently applied.

3.2.4. Reshaping translator roles and professional development (S6)

No category in the dataset reveals a sharper regional divide than this one. In the aggregate, 29.2% of papers (14 of 48) address the reshaping of translator roles. But the aggregate hides the real story. In the Chinese literature, 40.6% of papers (13 of 32) engage with this topic at length, treating the professional identity of the translator as a central pedagogical concern. In the English literature, just one paper (6.2%) mentions it. What Chinese scholars describe is a fundamental redefinition of what a translator is: not a person who converts text between languages, but a comprehensive language service provider, a project manager, or a "translation engineer" whose expertise includes technology integration as much as linguistic skill. The framing is macro-level and institutional: the argument is that educational programs have a responsibility to anticipate the professional landscape their graduates will enter and to align their training accordingly. Li [9] provides a competency map for this redefined professional identity, while Zhu and Wang [10] address the epistemological questions that arise when translation is reconceived as a technology-mediated service industry rather than a linguistic craft.

3.2.5. Reconstruction of translation quality assessment systems (S3)

A smaller but persistent thread in the literature concerns assessment. The question is what to do about grading when students can produce translations that look polished but may not be their own work, or that may contain AI-generated errors the student did not catch. Ten papers in the sample (20.8%) engage with this, and the English-language side shows proportionally more interest—31.2% versus 15.6% in the Chinese literature. The proposed reforms share a direction: away from grading only the final product and toward evaluating the process that produced it. Several studies sketch what they call a dual-track system. On one track, AI handles the mechanical checks that it does reasonably well—grammar, terminology consistency, formatting. On the other track, a human instructor evaluates the dimensions where context and judgment are irreducible: cultural appropriateness, creative decision-making, and the ethical choices embedded in how a text has been rendered. Lu and Feng [11] and Zhang and Peng [12] illustrate how such systems could work, though both acknowledge a current imbalance: the AI track is serviceable at flagging problems but weak at explaining why something is a problem, which limits its standalone pedagogical value.

3.2.6. Cultivation of cultural transfer competence (S4)

One paper. Out of 48. The English sample contains zero papers on this topic. The Chinese sample contains one (3.1%). That is the entire research base for how to teach students to handle the cultural dimensions of AI-assisted translation. Given the frequency with which the papers in this sample cite examples of culturally inappropriate or insensitive AI translations, the mismatch between the recognized problem and the research response is difficult to explain—except perhaps by noting that cultural transfer is inherently harder to systematize than feedback mechanisms or assessment rubrics, and thus harder to study with the methods that currently dominate the field. Wang and Liu [18] offer the only dedicated treatment, and it is explicitly a preliminary framework that calls for substantially more work. If cultural competence is as central to professional translation as the field claims it is, then this gap deserves considerably more research attention than it has received.

4. Discussion

Taken together, the quantitative findings and the cross-community comparisons reveal a field in which the challenges are unevenly distributed and the responses reflect deeper assumptions about how education should adapt to technological change. Three dynamics from the data merit closer examination.

4.1. Why does feedback failure drive curriculum reform

The data suggest a causal thread running from the most frequent challenge—the breakdown of feedback mechanisms (P6)—to the most frequent response—curriculum and pedagogical innovation (S5). Good translation teaching has always depended on detailed, context-sensitive feedback from instructors who can see not just what a student got wrong but why, and who can communicate that reasoning in terms the student can absorb. When GenAI floods the instructional environment with surface-level, decontextualized feedback, the traditional feedback loop breaks. This is not a problem that can be solved at the margins—by teaching students to ignore bad AI suggestions, say, or by adding an extra round of instructor review. Because the feedback mechanism is the central channel

through which translation competence is developed, its failure forces a more fundamental rethinking of how translation is taught.

The intensity of the curriculum reform response is higher in the Chinese literature (93.8%) than in the English literature (62.5%), and this gap is probably not coincidental. Chinese higher education operates within a more centralized institutional framework, where technological disruption tends to trigger coordinated, system-level responses rather than piecemeal adjustments by individual instructors. When a problem as pervasive as feedback failure is identified, the institutional machinery mobilizes around curriculum-level solutions. Jiao et al. [4] and Hu and Li [19] both trace the path from feedback inadequacy to the need for process-oriented evaluation, suggesting that the connection is not merely statistical but reflects a genuine pedagogical logic: when you cannot fix the feedback channel, you have to redesign the learning pathway that the feedback was supposed to support.

4.2. Is human-AI collaboration a solution or a new risk

Here is a tension that runs through enough of the literature to warrant its own heading. Structured human-AI collaboration (S2, supported by 39.6% of papers) is the strategy most frequently proposed to stop translation competence from eroding (P3, flagged in 45.8% of papers). But collaboration, if it means letting AI take the first pass at a translation problem, can itself become a vector for the very erosion it is meant to halt. The dynamic is not hard to trace. When a student reaches for ChatGPT before doing their own analytical work—before identifying the hard parts of the source text and making provisional decisions about how to handle them—the cognitive effort that builds translation skill gets replaced by the cognitive effort of evaluating AI output. Those are different skills, and strengthening the second does not compensate for weakening the first.

This is not an argument against collaborative models, which almost certainly represent the direction the field needs to move. But it is an argument for building guardrails into those models. Requiring students to complete an independent source-text analysis before turning to AI, as several of the guided learning proposals suggest, is one such guardrail. Monitoring cognitive effort throughout the translation process—rather than evaluating only the final output—is another. Freeth and Toto [20] make the case for an action research methodology that tracks how cognitive engagement shifts when AI tools are introduced, and they find that the timing of AI intervention matters considerably. Wang and Zhang [17], working from a Chinese institutional context, emphasize the design of explicit collaboration protocols. Despite the different research traditions, both sets of authors converge on the same practical insight: collaboration without structure is likely to do more harm than good.

4.3. Who should lead: the learner or the institution

Perhaps the most revealing divergence in the dataset concerns who is expected to drive the adaptation to GenAI. The Chinese literature leans heavily toward institution-led role reshaping (S6, 40.6%), treating the redefinition of the translator's professional identity as a task for educational programs, industry bodies, and policy frameworks. The English literature barely addresses this dimension (6.2%), focusing instead on building individual translation literacy (S1, 43.8%)—that is, on equipping learners with the critical awareness and independent problem-solving skills they need to navigate a technologically transformed profession on their own terms.

These are not simply two alternative strategies for achieving the same goal. They rest on different assumptions about where agency resides in educational change—in the structures that shape training

programs or in the capacities of individual learners. Zhang and Doherty [2] articulate a learner-centered view in which the primary task is to develop students' ability to think critically about AI outputs and to make informed decisions about when and how to use technological tools. Li [9] and Zhu and Wang [10], by contrast, frame the challenge in structural terms, arguing that the translation profession is being reshaped by forces that individual learners cannot control and that educational institutions have a responsibility to align their programs with those forces proactively.

Both perspectives have practical value, and there is no reason in principle why they cannot coexist. But the gap between them matters because it shapes what gets researched, what gets funded, and ultimately what translation classrooms look like. A program designed around the assumption that adaptation is an individual learner's responsibility will make different choices about curriculum structure, assessment design, and faculty development than one designed around the assumption that institutions should lead the transformation. The field would benefit from more explicit engagement with this underlying philosophical question.

5. Conclusion

This review has surveyed the relevant literature drawn from CNKI core journals and the Web of Science core collection, synthesizing with some quantitative grounding the challenges GenAI has introduced into written translation education and the pedagogical pathways proposed in response. Several limitations should be noted. The search was confined to two core databases and did not extend to publications outside those collections—monograph chapters, conference proceedings, and work published in other languages may well contain relevant material not captured here. In addition, the coding was carried out by a single researcher without an inter-coder reliability check, which means a degree of subjective judgment is embedded in the classification.

A number of directions follow from the findings. Whether structured human-AI collaborative pedagogies can genuinely sustain translation competence over time—rather than merely shifting the conditions under which skills erode—calls for longitudinal work of a kind the field currently lacks. Most of the evidence now available is cross-sectional, and for a phenomenon evolving at this pace, that base is thinner than it needs to be. The teaching of cultural transfer competence in an AI-mediated environment has barely registered in the literature surveyed here, yet LLMs will keep producing culturally awkward or inappropriate output for the foreseeable future. Translation programmes that do not build cultural judgment into their curriculum design risk sending graduates into the profession without a fully rounded skill set. On a broader level, the persistent divergence between English and Chinese scholarship over translator role reshaping and learner agency deserves more than a footnote. Different starting assumptions lead to different curriculum architectures and assessment logics, and translation educators would benefit from bringing that premise out of the background and into direct discussion. At a moment when the speed of technological change routinely outpaces the cycle of pedagogical research, the framework of problems and strategies that this review has mapped out may offer a set of reference points for educators looking to steady their course.

Funding project

This work was supported by the Chinese Society of Educational Development Strategy - General Project (No. ZLB20240832), titled "Research on the Impact of ChatGPT on Computer-Assisted Translation Education Model and Countermeasures," and the Hubei Provincial Education Science Planning - Key Project (No. 2025GA141), titled "Research on the Innovation of Translation

Classroom Teaching Mode and the Improvement of Learning Ability Based on Large Language Models."

References

[1] Wang, H., & Xie, F. (2024). Research on the innovation of translation education practice models driven by large language model technology. Chinese Translators Journal, 45(2), 70-78.
[2] Zhang, J., & Doherty, S. (2025). Investigating novice translation students' AI literacy in translation education. The Interpreter and Translator Trainer, 19(3-4), 234-253.
[3] Hu, K. (2026). Transformation and development of translation disciplines in China under the background of artificial intelligence. Shandong Foreign Language Teaching, 47(1), 2-12.
[4] Jiao, H., Hu, W., & Zhang, X. (2025). To eat or to feed: Can large language models provide useful feedback in translation education? The Interpreter and Translator Trainer, 19(3-4), 317-337.
[5] Zhang, L., Li, L., Jiang, J., et al. (2026). Exploring the impact of diverse feedback sources on learners' performance, motivation, and preference in a translation course: Tutor, peer, and GPT insight. Thinking Skills and Creativity, 59, 102042.
[6] Zhang, W. (2024). Challenges and solutions for translation programs in the AI era: An analysis based on a large-scale social survey. Chinese Translators Journal, 45(5), 139-148.
[7] Geng, F., & Hu, J. (2023). New directions in AI-assisted post-editing: A case study based on ChatGPT. Foreign Languages in China, 20(3), 41-47.
[8] Yao, Y., Han, T., & Li, D. (2025). Measuring translation trainees' effort in AI-assisted post-editing: A multi-method approach. The Interpreter and Translator Trainer, 19(3-4), 357-378.
[9] Li, Z. (2024). Translator professional development in the digital-intelligent era: Situation, competencies, and training pathways. Technology Enhanced Foreign Language Education, (6), 8-14.
[10] Zhu, Y., & Wang, L. (2025). The epistemology of translation technology in the generative AI era: Theoretical connotations and pedagogical implications. Chinese Translators Journal, 46(1), 62-70.
[11] Lu, X., & Feng, L. (2024). Application of AI-empowered CSE written translation competence scale in teaching evaluation. Foreign Language World, (6), 29-36.
[12] Zhang, J., & Peng, S. (2025). An empirical study on LLM-empowered intelligent assessment of student translations. Contemporary Foreign Language Studies, (5), 85-96.
[13] Mu, Y., & Zheng, S. (2025). Construction of AI-empowered translation knowledge network and new paradigm of human-machine collaborative teaching. Technology Enhanced Foreign Language Education, (3), 52-61.
[14] Wang, H., Wang, S., & Wang, W. (2026). The future of intelligent translation: How large language models reshape translation education. Shandong Foreign Language Teaching, 47(2), 1-9.
[15] Wang, S., & Luo, X. (2025). Intelligent translation literacy in the GenAI era: Practical foundations, academic rationale, and conceptual framework. Foreign Language Education, 46(1), 59-65.
[16] Tian, S., Wang, D., Wang, J., et al. (2025). Empowering GenAI with a guidance-based approach in MTPE learning: Effect on student translators' cognitive process, final translation quality and learning motivation. The Interpreter and Translator Trainer, 19(3-4), 379-404.
[17] Wang, J., & Zhang, Z. (2025). Exploring human-machine collaborative translation teaching models driven by large language models. Foreign Language World, (5), 69-76.
[18] Wang, H., & Liu, S. (2025). From MTPE to AIPE: The evolution of translation models in the GenAI era and its implications for translation education. Shandong Foreign Language Teaching, 46(3), 111-121.
[19] Hu, K., & Li, J. (2024). Translation talent cultivation in the context of large language models: Challenges and prospects. Technology Enhanced Foreign Language Education, (6), 3-7.
[20] Freeth, P. J., & Toto, P. (2026). The impact of AI-in-the-loop pedagogy: An action research approach to integrating AI technologies in translator training. The Interpreter and Translator Trainer, 1-18.