Asian Journal of University Education (AJUE)
Volume 22, Number 2, June 2026

Effectiveness of AI Translation Tools on EFL Translation Skills in Higher Education: A Systematic Review

Yi Man[1], Pavani Malaa Meganathan[2]*

[1]Faculty of Business, Information & Human Sciences, Kuala Lumpur University of Science and Technology,
Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia.
232924055@s.klust.edu.my
[2]Faculty of Modern Languages and Communication, Universiti Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia.
pavani@upm.edu.my
*Corresponding Author

https://doi.org/10.24191/ajue.v22i2.010

Received: 5 February 2026
Accepted: 15 February 2026
Date Published Online: 25 May 2026
Published: 25 May 2026

Abstract: This systematic review investigates how AI translation tools support EFL learners' translation learning by synthesizing empirical evidence on learner strategies, effectiveness, and pedagogical implications. Following PRISMA-informed procedures, 26 peer-reviewed studies published between 2008 and 2025 were screened and analyzed. To enable systematic cross-study comparison, the evidence was organized using an Environment-Task-Learner-Strategy (ETLS) framework. The synthesis shows that AI-assisted translation environments are predominantly shaped by readily accessible tools, with neural machine translation systems (e.g., Google Translate) dominating earlier studies and increasing attention to large language model tools (e.g., ChatGPT) in recent work. Across contexts, learners demonstrate a recurring repertoire of strategies, including cross-tool comparison and verification, reverse translation checks, selective tool use for difficult segments, iterative refinement (including prompt adjustment in LLM settings), and systematic post-editing. With respect to effectiveness, the evidence converges on short-term performance benefits such as improved linguistic accuracy and fluency, while also indicating boundary conditions. Learning benefits are more consistently observed when learners critically evaluate and revise AI-generated output rather than adopt it uncritically, and when outcomes are assessed using translation-relevant criteria. The findings emphasize the importance of structured AI translation literacy, post-editing–oriented instruction, proficiency-sensitive support, and more rigorous research designs to assess sustained translation competence.

Keywords: AI translation tools; EFL translation skills; large language models; learner strategies; machine translation

1. Introduction

AI-powered translation technologies are reshaping both professional translation and educational practice, as manual translation alone can no longer meet the speed, scale, and quality demands of contemporary multilingual communication (Koka, 2024). With the rapid expansion of generative AI, there is growing urgency to examine how AI translation tools function as learning technologies that is what they enable, what they constrain, and under what conditions they support measurable learning in translation tasks (Salloum et al., 2024).

Existing evidence on machine-translation-assisted language learning (MTALL) suggests that machine translation (MT) can contribute to language development and classroom integration, yet findings remain dispersed across tools, instructional settings, task types, and outcome operationalizations, limiting cross-study comparability and cumulative conclusions (Jiang et al., 2024). Moreover, much of the synthesis has centered on MT within broader second language (L2) domains (e.g., writing or general mediation), rather than treating translation-skill development as a distinct construct with translation-relevant criteria. For tertiary translation education, 'effectiveness' should be demonstrated through translation-specific indicators (e.g., adequacy/accuracy, appropriateness, error patterns, culturally sensitive meaning construction), not only generic language achievement or attitude measures (Kirchhoff, 2024; Deng & Yu, 2022). A related gap concerns learner regulation and strategy use in AI-assisted translation. While learners frequently report positive perceptions and extensive MT use, learning benefits appear heterogeneous and shaped by proficiency and usage patterns; recent evidence also cautions that self-directed MT use may sometimes encourage L2 avoidance, suggesting that access to high-quality AI output does not automatically translate into deeper processing or transferable competence (Murtisari et al., 2024). However, strategy-level evidence remains insufficiently consolidated regarding how EFL learners combine cognitive and metacognitive regulation with emerging AI-interaction strategies (e.g., prompting, iterative refinement, cross-tool comparison, verification), and what forms of instructional scaffolding promote critical engagement rather than overreliance (Dinh, 2025).

Since late 2022, generative AI has further accelerated interest in AI-supported language learning. Reviews of ChatGPT in language education describe a rapidly expanding but methodologically uneven evidence base and highlight the need for stronger research designs and outcome measures beyond perception-driven claims (Li et al., 2024). Although EFL focused studies suggest potential benefits when ChatGPT is integrated as instructional support, translation-specific evidence remains fragmented across tool types, genres, and evaluation frameworks (Teng, 2024). Recent comparative evaluations also indicate that Large Language Model (LLM) performance relative to mainstream MT systems depends on prompting conditions and culturally loaded discourse demands-precisely the boundary conditions that translation pedagogy must address (Chen & Lin, 2025). Taken together, these developments justify a systematic synthesis that clarifies what has been done, what the evidence indicates about effectiveness, and how AI-assisted translation learning in EFL contexts is likely to evolve.

Guided by PRISMA reporting principles, this systematic review synthesizes empirical evidence on the effectiveness of AI translation tools for developing EFL learners' translation skills by: (1) identifying and classifying learner strategies when engaging with AI-generated translations; (2) evaluating reported effectiveness of AI translation tools for improving translation performance outcomes; and (3) synthesizing evidence on AI translation tools in EFL translation learning to derive concise implications for future research, pedagogical implementation and classroom practice.

2. Literature review

Research on AI for translation has evolved alongside major paradigm shifts in machine translation. Early conceptualizations framed automatic translation as an AI problem, drawing on cryptography and information theory (Reifler & Pol, 1962). Landmark milestones include the Georgetown-IBM demonstration in 1954, often cited as the first large-scale public MT showcase (Dostert, 1955; Hutchins, 2005). After decades dominated by rule-based approaches, statistical machine translation (SMT) became the leading paradigm in the 1990s, leveraging probabilistic modeling trained on bilingual corpora (Brown et al., 1993; Koehn, 2009). The mid-2010s then saw the rise of neural machine translation (NMT), where sequence-to-sequence and attention-based architectures produced major gains in fluency and adequacy (Sutskever et al., 2014; Bahdanau et al., 2015). Since around 2016, commercial MT systems have widely deployed NMT (e.g., Google's GNMT), forming the technological backbone of many AI translation tools now used in educational settings (Wu et al., 2016).

Within EFL education, AI translation tool use has become pervasive, expanding from informal learner support to increasingly structured pedagogical integration. As free online translators became widely accessible in the late 2000s, learners began using MT to manage linguistic difficulty, often adopting selective "last resort" patterns, attempting to produce as much English as possible and consulting MT for segments beyond their proficiency (García & Peña, 2011). As tool quality improved in the 2010s, reliance expanded and use became embedded in everyday study routines beyond the classroom, reflecting the ubiquity of mobile and web-based translators. This normalization has also reshaped pedagogical debates: rather than treating MT as a prohibited shortcut, recent work increasingly focuses on how to integrate AI translation tools in ways that support learning, with instructional attention shifting from whether learners use MT to how they use it and how teaching can scaffold critical engagement (Kirchhoff, 2024). Consequently, many contemporary studies examine planned classroom interventions or training activities, while others continue to document spontaneous out-of-class use and learner perceptions, together indicating that AI translation tools are now a mainstream resource requiring evidence-informed pedagogical guidance.

Empirically, the literature reflects diverse educational contexts but is concentrated in higher education. Most studies are situated in university EFL settings, including general English, academic writing, and translation-focused courses, with additional evidence emerging from secondary schooling and adult education, suggesting diffusion beyond tertiary contexts (Cancino & Panes, 2021; Alghasab, 2025). Research most commonly examines widely available public MT systems, especially Google Translate, often alongside comparative tools (e.g., DeepL or Bing Translator). More recently, generative AI systems such as ChatGPT have been explored as translation supports in classroom scenarios, reflecting rapid uptake of new AI capabilities (Abdelhalim et al., 2025; Dinh, 2025). Implementation typically relies on accessible web/mobile platforms rather than bespoke software, aligning study contexts with authentic learner practices. Only a minority of studies incorporate dedicated computer-assisted translation (CAT) environments; for example, Li and Tekwa (2025) report integrating an online CAT platform with MT functionality and instructor feedback in a semester-long course. Overall, current trends suggest that EFL translation education is increasingly shaped by readily available AI translation technologies, making systematic synthesis essential for understanding how tool type, instructional design, and learner engagement patterns jointly influence translation learning outcomes.

Although research on AI-assisted translation in EFL contexts is expanding, the evidence base is still constrained by intertwined conceptual and methodological limitations. Conceptually, many studies do not operationalize translation competence with sufficient precision and sometimes treat improvements in general L2 writing fluency or accuracy as proxies for translation-skill development, leaving translation-relevant constructs (e.g., meaning fidelity, adequacy, genre/pragmatic appropriateness, and cultural mediation) under-specified and difficult to compare across studies (Lee, 2020; Yang et al., 2023). In addition, the literature often conflates performance convenience with learning: higher-quality AI outputs may reflect tool-enabled production rather than durable competence gains, particularly when learners adopt AI suggestions uncritically or use tools to bypass cognitively demanding processing (Kirchhoff, 2024; Deng & Yu, 2022). A further gap is the weak integration of learning theory and strategy frameworks; learner behaviors (e.g., post-editing, reverse checks, cross-tool comparison, iterative prompting) are frequently reported descriptively but not consistently categorized or theoretically interpreted, limiting cumulative insight into how strategic engagement mediates learning outcomes (García & Peña, 2011; Murtisari et al., 2024; Dinh, 2025).

Methodologically, studies rely heavily on perception-based evidence and employ heterogeneous outcome measures (e.g., Pikhart et al., 2024; Dinh, 2025), with limited use of standardized translation-performance metrics, controlled comparisons, or baseline/post designs that can isolate tool effects (Li et al., 2024). The empirical base is also dominated by small, single-context, short-duration investigations (such as Lee, 2020; O'Neill, 2019), and substantial variation in research designs and tool/task configurations further reduces generalizability and complicates cross-study synthesis (Kirchhoff, 2024).

2.1 The ETLS Framework

To address these conceptual ambiguities and methodological inconsistencies, this review adopts a structured ecological synthesis framework. This study adopts an AI-assisted translation learning ecology framework, operationalized as ETLS (Environment- Task- Learner- Strategy), to structure both the synthesis of evidence and the positioning of the present inquiry.

Within ETLS, Environment refers to the material and instructional conditions under which AI tools are deployed for translation learning, including the type and affordances of AI translation tools (e.g., neural machine translation, generative AI assistants, CAT-like functions), the learning setting (classroom-based vs. blended; mobile vs. desktop). Task captures the design features of translation learning activities, such as directionality (E-C/C-E), genre and domain specificity (general vs. business/ESP), unit of translation (sentence/paragraph/text), and assessment expectations (rubric-based accuracy, fluency, terminological consistency, pragmatic and cultural appropriateness), as well as the extent to which tasks require post-editing, comparison across outputs, and justification of translation choices. Learner highlights the characteristics of EFL students that may condition outcomes, including baseline language and translation competence, digital/AI literacy, motivational profiles, and technology acceptance tendencies. Strategy focuses on the observable and reported patterns of translation-related strategy use in AI-mediated contexts, spanning cognitive strategies (e.g., segmentation, reformulation, terminology verification), metacognitive strategies (e.g., monitoring, evaluating, reflecting), and AI-specific interaction strategies (e.g., prompt refinement, triangulation with parallel texts/dictionaries, systematic post-editing and error diagnosis).

Building on this ETLS framework, this study formulates the following research questions to guide the literature synthesis:

RQ1. What are the AI-assisted translation strategies do EFL students report using in translation learning contexts?
RQ2. What does the existing evidence suggest about the effectiveness of AI translation tools for improving EFL learners' translation performance?
RQ3. What implications does this evidence offer for future research, pedagogical implementation, and the evolving role of AI in EFL translation learning?

3. Methodology

Given the heterogeneity in how included studies operationalized 'translation competence', this review adopted a functional operational definition to enable cross-study comparison. For the purposes of synthesis, translation competence was defined as measurable performance on translation-specific tasks involving meaning transfer between languages, evaluated through criteria such as accuracy/adequacy, fluency, lexical and syntactic appropriateness, error patterns, and cultural or pragmatic sensitivity.

Studies that assessed general L2 writing fluency or complexity were included only when outcomes were explicitly derived from translation tasks or when writing measures were directly tied to translation performance. Where studies treated general linguistic improvement as a proxy for translation competence, this review interpreted such findings cautiously and categorized them as indirect indicators rather than direct measures of translation-skill development.

During synthesis, outcome measures were therefore grouped into translation-specific performance indicators and general language development measures associated with translation tasks. Greater interpretive weight was given to studies employing explicit translation evaluation rubrics or task-based assessment aligned with translation constructs.

3.1 Literature Search Strategy

The following scholarly databases were systematically searched for peer-reviewed journal articles: Education Resources Information Center (ERIC), Web of Science, ScienceDirect, Scopus and Education Research Complete. In addition, Google Scholar is used as a supplementary search tool to uncover any relevant studies not indexed in the primary databases.

To capture the multifaceted topic, three thematic keyword groups have been formulated based on the core concepts of the study. These groups were linked with the Boolean operator AND to narrow results to studies covering all relevant aspects. Each group contains polished English terms (including synonyms and related phrases) reflecting academic terminology in AI-assisted language learning, EFL translation training and assessment. Table 1 details the search terms used across the three domains.

Table 1

The Search Terms

Domain	Search Term
Translation Skills	"translation" or "translator" or "translation competence" or "translation proficiency" or "translation strategies" or "translation training" or "translation assessment" or "translation evaluation" or "translating"
AI-Assisted Learning	AI-Assisted Learning ("AI" or "machine" or "AI translation tools" or "AI tools" or "Artificial Intelligence" or "ChatGPT" or "computer" or "machine translation" or "AI-assisted language learning" or "computer-assisted" or "AI integration" or "neural machine translation" or "GenAI"
EFL Context	"EFL" or "Foreign Language" or "Second Language" or "L2 learners"

This keyword strategy ensures that search queries target literature at the intersection of translation pedagogy, AI tool usage and EFL education. By using an "AND" linking strategy between these conceptual groups (as modeled in the reference review), the search will filter for studies that simultaneously address all key dimensions of the topic.

3.2 Inclusion and Exclusion Criteria

Clear inclusion and exclusion criteria guided the selection of studies from the search results. These criteria were applied during title/abstract and full-text screening to ensure that included studies directly address how AI translation tools influence EFL learners' English translation skill development.

The literature search is restricted to the period 2008-2025. The selected time window therefore captures both the initial integration of machine translation into EFL/ESL teaching and the more recent developments associated with neural machine translation and AI-assisted learning environments. Within this range, the search is limited to English-language, peer-reviewed journal articles, ensuring scholarly quality and accessibility. Particular attention were given to studies published in the last 5 years (2020-2025), when neural MT and classroom-oriented AI applications have become especially prevalent, so that the synthesis reflects the most current technological and pedagogical developments.

Table 2

Inclusion and Exclusion Criteria

Inclusion Criteria (must meet all)	Exclusion Criteria (excluded if any apply)
● Participants are EFL learners. ● Educational or training-oriented translation context (e.g., vocational education, colleges/universities, adult education, professional/translator training). ● Examines AI tools used for translation tasks, including (but not limited to) machine translation software, AI-based translation apps, or AI writing assistants used specifically for translation-related work. Includes classroom studies, curriculum interventions, and comparative designs (AI-assisted vs. traditional instruction). ● Reports English translation competence/skills as a measurable outcome (e.g., translation accuracy,	● Studies focusing exclusively on K-12 or younger learners (primary/secondary). ● Non-educational/non-learning use only (e.g., travel, casual consumer use, workplace communication) without explicit learning aims, instructional design, or translation-skill outcomes. ● Studies using AI/CALL tools without a translation component (e.g., general writing tools with no translation tasks or translation-related outcomes). ● Does not measure translation-skill outcomes (e.g., focuses only on attitudes, general writing improvement, or language outcomes unrelated to translation competence).

Inclusion Criteria (must meet all)	Exclusion Criteria (excluded if any apply)
translation quality evaluation, test/assessment-based proficiency gains, or comparable metrics). ● English is a working language, and the study is situated in an EFL translation context.	● Translation pedagogy in language pairs not involving English (e.g., Spanish–French without an English component). ● Focuses on outdated MT/pre-modern AI (e.g., early rule-based systems) that do not represent current AI capabilities. ● Non–peer-reviewed outputs (e.g., theses, conference abstracts without full papers, reports) and/or publications outside 2008–2025, and/or non-English publications.

3.3 Screening Methodology

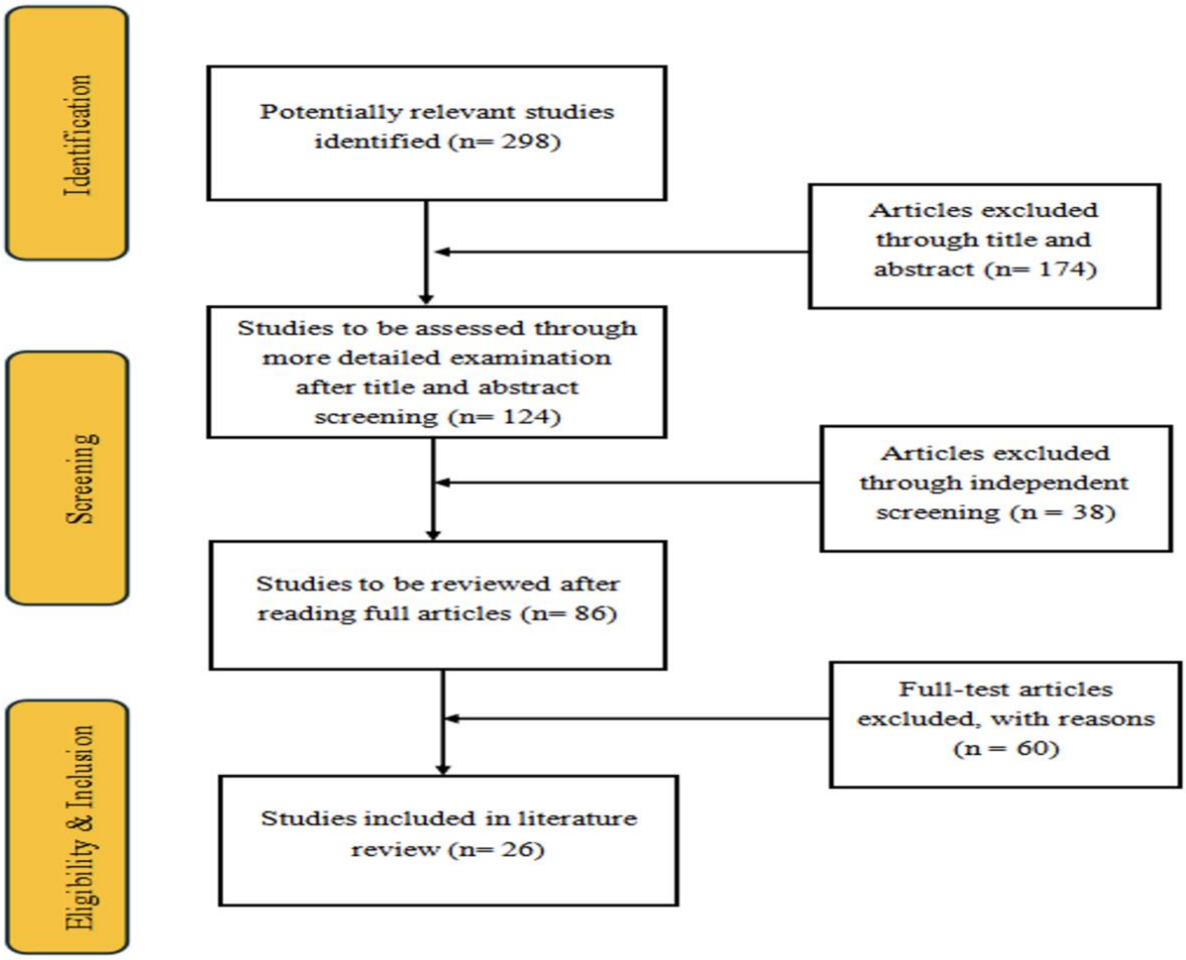
A rigorous screening process was followed, modeled on best practices for systematic literature reviews (e.g., PRISMA guidelines). The screening proceeded in multiple stages to filter the initially retrieved articles down to those that truly meet the criteria:

i. Identification: All references retrieved from the database searches were imported into a reference management tool (or spreadsheet). This initial pool might contain hundreds of records (as seen in the reference study, which initially identified 298 articles after de-duplication). Duplicate entries (the same article found in multiple databases) were removed at this stage to yield a unique list of studies.
ii. Title and Abstract Screening: Each remaining study's title and abstract were reviewed against the inclusion/exclusion criteria. This is a preliminary relevance check.
iii. Full-Text Review: For studies that pass the abstract screening, full texts were obtained and read in detail. Using a standardized form or checklist, each article was assessed to confirm it meets all inclusion criteria and does not meet any exclusion factors upon closer inspection. During this stage, reasons for exclusion were logged to maintain a transparent record. Any uncertainties or borderline cases were discussed among the research team. This step yields the final set of articles were included in the review.
iv. Quality Appraisal: Alongside the full-text review, a quality assessment of each study were conducted. This involves evaluating the rigor and trustworthiness of the research.
v. Data Extraction and Synthesis: (Moving beyond screening to the analysis phase, if applicable). Key data from each included study was extracted - e.g., study context, sample, details of the AI tool used, study design, and main findings regarding translation skill outcomes. This facilitated a structured comparison of studies in the eventual write-up. A PRISMA flow diagram was prepared to document the number of studies at each stage of screening (identified, screened, eligible, included) and reasons for exclusions, ensuring transparency in the review process as per international standards (Moher et al., 2009).

After screening and full-text analysis, 26 studies met all inclusion criteria and were included in the final synthesis. Figure 1. displays the flow diagram of article search and screening steps.

Figure 1

Flow diagram of article search and screening steps to identify studies for inclusion in the review



Following established systematic review practices, a detailed coding form was developed to extract key information from each study. For each article, the researchers recorded its reference details, research design and methods, and the main findings and conclusions. This study then catalogued specific data pertinent to this review's focus, including: (1) research design and methods; (2) the strategies students used in conjunction with AI translation (as described by the authors); (3) effectiveness of AI translation tools for improving translation performance outcomes; and (4) implications for future research, pedagogical implementation, and the evolving role of AI in EFL translation learning.

This review was conducted and reported in accordance with PRISMA 2020 guidelines. Using this coding scheme, the articles were abstracted systematically and compiled into a comprehensive summary table (see Appendix Table A. for an overview of all studies). In constructing this table, the present study preserved, as much as possible, the original language used by study authors to describe their findings and participants, to minimize interpretation bias.

3.4 Quality Appraisal and Risk of Bias

To enhance transparency and interpretive rigor, a light-touch quality appraisal was conducted for all included studies. Given the methodological diversity of the literature (including experimental, quasi-experimental, mixed-methods, and qualitative designs), a flexible appraisal approach was adopted rather than a single standardized risk-of-bias tool. Each study was evaluated against three broad criteria commonly used in educational systematic reviews: (1) clarity and appropriateness of research design,

(2) validity and relevance of translation-related outcome measures, and (3) transparency of data collection and analysis procedures.

Studies were not excluded on the basis of quality alone; instead, appraisal outcomes were used to contextualize findings during synthesis. In particular, greater interpretive weight was given to studies with clearly defined translation tasks, explicit evaluation criteria, and outcome measures aligned with translation competence, while findings from perception-only or methodologically limited studies were treated cautiously. Where relevant, potential sources of bias (e.g., reliance on self-report, short intervention duration, absence of comparison groups) are acknowledged in the Discussion.

4. Findings

4.1 RQ1. What Are the AI-Assisted Translation Strategies Do EFL Students Report Using in Translation Learning Contexts?

Nearly all the reviewed studies (with only a few exceptions where strategy use was not examined) reported insights into how EFL students interact with and exploit AI translation tools, revealing a repertoire of strategies. A first broad strategy observed is cross-referencing and verification. Students frequently did not take AI outputs at face value; instead, they engaged in cross-checking. For example, inputting the same text into multiple translation engines and comparing results, or using the reverse-translation technique (translating the AI's output back into the L1 to check fidelity). Vukanović (2025) noted that many of his Croatian students would run difficult sentences through both Google Translate and ChatGPT, and then reconcile the outputs, an approach of triangulating between tools. This practice of comparing outputs also ties to another common strategy: leveraging MT as a dictionary or reference tool rather than a one-shot translator. Especially for single words or short phrases, learners treated the MT as a bilingual dictionary, checking definitions or alternative translations for vocabulary items (Daniel & Ancheta, 2025). In the study by Daniel and Ancheta, such "GT-as-dictionary" use was one of the most prevalent strategies across Arab EFL learners. By examining multiple translation options, students attempt to ensure accuracy and choose the most appropriate wording, indicating a strategic approach to overcome the known limitations of any single AI output.

A second set of strategies revolves around iterative refinement and post-editing. Rather than using the AI tool in a single step, students often engaged in iterative cycles: they would edit the AI's output or adjust their input and run it again to improve the results. For instance, Dinh (2025) documented that Vietnamese EFL students, when using ChatGPT for translation, would refine their prompts (providing more context or instructions) if the initial output was unsatisfactory- a strategy of prompt tuning to guide the AI. Similarly, when using standard MT, students applied post-editing strategies: correcting obvious grammatical errors, adjusting word choice, and fixing word order in the machine's translation. Several studies explicitly measured how well students post-edited AI outputs. The general strategic behavior here is that learners do not view the AI output as final; they see it as a draft to be improved. This is echoed by self-reported strategies in surveys: for example, multiple participants in Pikhart et al. (2024) indicated that they "use Google Translate then polish the English result"- polishing meaning they review and edit the translation by themselves or with peers. This post-editing mindset is a positive sign, as it means learners are actively engaging with the language output, potentially learning from the corrections they must make.

An interesting specific strategy related to the above is input simplification or controlled writing to optimize MT output. Especially among more experienced users, there was awareness that the quality of machine translation can be influenced by how the source text is written. Chomicz (2025) observed that her adult learners intentionally wrote shorter, simpler sentences in Polish before feeding them to the translator app, to get a clearer English translation that they could then expand upon. Likewise, some students reported avoiding idioms or complex phrases in the input to prevent mistranslation. This kind of strategic adaptation- essentially pre-editing one's own L1 text- demonstrates metacognitive understanding of the AI's limitations and how to work around them. It aligns with strategies professional translators use with MT (writing in a "MT-friendly" way) and suggests that learners can develop similar savvy. Another manifestation of this was giving detailed prompts to ChatGPT: students discovered that by specifying the context or desired style (e.g., "Translate this passage into English using formal academic language"), they could steer the output closer to their target, rather than using the default output blindly.

A further prevalent strategy is selective use of AI translation within a task. Students do not necessarily translate entire texts with the tool; they selectively translate difficult sections or unfamiliar phrases, while handling easier parts themselves. García and Peña (2011) noted this behavior in beginner Spanish learners: students wrote whatever English sentences they could on their own, and when stuck, they used MT for the parts beyond their ability. In more recent studies, this pattern persists. For example, in Lee's (2020) study (replicated by Yang et al., 2023), many learners drafted an essay partly on their own and partly via MT- often writing the introduction and conclusion by themselves but resorting to MT for complex body paragraphs or vice versa. This hybrid strategy can yield a patchwork of human and machine text, which students then try to smooth out. It reflects a practical approach to maximize efficiency: using AI help where it provides a clear benefit (speeding up production of text they find challenging), but not for everything. Some instructors have described this as students using MT as a "last resort" or backup tool (O'Neill, 2019), which indicates a strategy of reserving the tool for specific challenges rather than defaulting to it for all writing.

Notably, language-switching strategies were also reported. Several Arab EFL learners in the Middle East studies described a pattern of drafting in Arabic then translating to English with GT, essentially using MT to convert their L1 output to L2 (Ali, 2022; Daniel & Ancheta, 2025). Conversely, a few studies noted learners checking their English output by translating it back into their L1 to see if the meaning stayed intact (a form of reverse validation). This back-and-forth translation use is a deliberate strategy to ensure comprehension and accuracy: by seeing how the machine renders the English text into their native language, learners can sometimes catch nuances or errors they missed. It's essentially a bilingual proofreading technique leveraging the MT in both directions.

Finally, an overarching strategic mindset observed among more successful learners is critical evaluation of AI outputs. Advanced students exhibited a cautious approach: they would scrutinize the AI's suggestions, accept what seemed useful, and reject or revise what seemed off target. Abdelhalim et al. (2025) report that their higher-level student translators, while preferring ChatGPT's more fluent translations, often double-checked facts and nuance, aware that the AI could be confidently wrong. In surveys, many learners mentioned "common sense" checks- reading the AI-produced translation to see if it logically and grammatically made sense in context before using it. Notably, Kruk and Kałużna (2025) report that even when students experienced gains in translation accuracy and motivation under AI/MT-supported conditions, some explicitly worried about overreliance and reduced engagement, reinforcing the need for verification and selective adoption as a pedagogical norm rather than an optional habit. The importance of this critical stance is emphasized in the literature: students are advised (and some intuitively do this) not to trust the AI blindly. For instance, one of the "key strategy" recommendations synthesized by Daniel and Ancheta (2025) is that learners should engage in verification and selective adoption, meaning they should verify the AI's output against other sources or their own knowledge and only adopt it into their work after ensuring it's correct. This can be considered a meta-strategy that governs all other strategies, ensuring that AI use remains a support for learning rather than a wholesale substitute for the learner's own thinking.

4.2 RQ2. What Does the Existing Evidence Suggest About the Effectiveness of AI Translation Tools for Improving EFL Learners' Translation Performance?

How effective are AI translation tools in improving EFL learners' translation abilities and related language skills? The studies in this review provide converging evidence that these tools can be beneficial under the right conditions, yet their impact is nuanced and not uniformly positive. Broadly, the effectiveness can be considered from two angles: performance outcomes (quality of translations/writing, test scores, etc.) and learner perceptions or attitudes. We discuss both, along with moderating factors that influence effectiveness.

On measures of performance and task outcomes, many studies reported improvements in students' work when AI tools were used, or at least no harm compared to traditional methods. For instance, in controlled experiments, students who had access to MT or ChatGPT often produced writing or translations with higher immediate accuracy and complexity. Particularly, ChatGPT assistance led to essays that were longer and richer in vocabulary and sentence structure, though sometimes at the expense of textual cohesion. In translation-specific tasks, AI tools likewise boosted performance on certain metrics. Abdelhalim et al. (2025) noted that both novice and advanced student translators produced more fluent target texts with ChatGPT's help than on their own, aligning with the tool's strength in fluency. Several corpus-based evaluations lend objective support: for example, Kwok et al. (2025) showed that student translations post-edited with GPT had higher lexical complexity (greater variety of word choice) than the students' original translations. These findings suggest that AI tools can raise the linguistic quality of output, providing a form of on-demand enhancement or correction that learners alone might not achieve.

However, the improvements in product do not always straightforwardly translate to improvements in underlying competence. A recurring theme is that while AI assistance makes the immediate task easier or the immediate product better, concerns remain about whether students are improving their language proficiency and translation skills in the long run. Some studies attempted to measure learning gains more directly. For example, Zhang et al. (2025) and Yang et al. (2023) looked at whether using AI tools over a semester improved students' independent writing or speaking skills. In Zhang et al.'s experiment with different NLP tools, they found all groups improved their English speaking slightly over 12 weeks, but no group (including the one using a chatbot translator) outperformed the control significantly, implying that AI tools did not drastically accelerate skill development, although they did aid task performance during the intervention. In Yang et al.'s replication study on writing, students who used MT for draft writing did not show a significant difference in post-test writing proficiency compared to those who wrote without MT, despite the MT group's drafts being better in the moment. These outcomes hint at a potential trade-off: AI tools can mask difficulties and produce a high-quality text for the learner, which might reduce the struggle that often leads to learning. This idea was explicitly raised by García and Peña (2011), who described the effect of MT as "positive for output quantity/quality, but possibly at the expense of learners' cognitive engagement". In their study, beginners who used MT wrote more and with fewer errors, but the authors cautioned that these students might not be internalizing all the corrections provided by the machine, as they bypass the trial-and-error process of composing in L2.

The perceptions and attitudinal data uniformly indicate that students value AI translation tools for the support they provide. In surveys and interviews across different contexts, a large majority of learners reported positive opinions, believing that tools like Google Translate or ChatGPT help them learn vocabulary, check their writing, and save time (Dinh, 2025; Pikhart et al., 2024; Ali, 2022). Motivation can also be positively impacted: for example, Bekdaş (2025) reported that integrating ChatGPT-based activities in a Turkish university EFL class led to a significant boost in student motivation, with students particularly appreciating the immediate feedback and "smart" assistance the AI provided. Advantages commonly cited by students include the speed of obtaining translations, improvements in spelling/ grammar, and the confidence of having a "second pair of eyes" on their writing. Notably, even teachers in some studies acknowledged benefits- Alghasab (2025) found that instructors, while cautious, saw potential in AI tools to assist students' writing process and encouraged their guided use.

Yet, reservations and concerns were also clearly documented, both by students and researchers. Chief among these is the issue of translation accuracy and errors: students are aware that AI translations can be wrong or unnatural, especially for nuanced or complex language. Abdelhalim et al. (2025) noted incidents of mistranslations or factual errors from ChatGPT (which can fabricate information), leading some learners to distrust the tool for high stakes use. The lack of contextual understanding was flagged- MT might choose an incorrect word sense or tone because it doesn't fully grasp the context or pragmatics, requiring the student to notice and fix it. Another concern is over-reliance: both educators and learners worry that easy access to AI translation might reduce learners' incentive to try using their own language skills. If a student habitually just pastes text into Google Translate without attempting to recall vocabulary or apply grammar rules, their productive skills might stagnate. A few studies provided evidence for this: for instance, a case in Chon et al. (2021) showed that some students who could have written a decent paragraph by themselves chose to translate an L1 paragraph with MT, resulting in correct output but possibly less practice of forming sentences in English. In interviews by Dinh (2025), some students themselves noted this dilemma: using ChatGPT made writing "too easy", and they feared they weren't practicing their English as much. This aligns with broader pedagogical concerns raised in the literature that uncritical use of MT can become a "crutch" and hinder the development of autonomous language production skills.

Importantly, many studies concluded that effective learning with AI tools is contingent on guidance and strategy. When students received training on how to use the tools (e.g., how to post-edit, how to interpret AI output critically, when to use or not use the translator), the outcomes were generally more positive. Students in Pikhart et al.'s (2024) study who were more satisfied with MT were often those whose teachers integrated it into lessons with clear goals, rather than those using it ad hoc. This trend supports the idea that pedagogical integration, incorporating AI translators as a tool to be skillfully used rather than a forbidden shortcut, leads to better learning. Daniel and Ancheta's (2025) comprehensive review echoes this: they found that while Google Translate is widely used and can support performance on routine tasks, its successful use requires learners to engage in critical post-editing and not to overly depend on it. They caution that without developing students' own language and translation competence, MT output alone does not automatically translate into learning gains.

Another moderating factor in effectiveness is learner proficiency, as mentioned earlier. Higher-proficiency learners often benefit more from AI tools because they can use them as a complement to their skills- they understand the AI's limitations and learn from its suggestions, whereas lower-proficiency learners might either misuse the tool or fail to correct its mistakes. For instance, Chung (2020) found that advanced Korean EFL students improved the clarity and accuracy of MT-translated texts more effectively than intermediate students during post-editing, meaning the advanced group gained more from the MT-assisted activity. This suggests that a baseline of L2 knowledge is needed to turn MT into a learning opportunity; otherwise, learners might accept erroneous output or not grasp why the MT's translation is the way it is.

In summary, the effectiveness of AI translation tools on EFL learners' translation skills is moderately to strongly positive when used strategically, and generally not detrimental to learning outcomes if over-reliance is avoided. Students improve their translated products and can learn from AI input, especially in terms of vocabulary and form, though the translation tools are not a magic bullet for proficiency. The tools work best as assistants- speeding up translations, providing suggestions and corrections, within a learning process still orchestrated by informed pedagogical practices and active learner engagement. This balanced perspective resonates with the findings across the reviewed studies: AI translation is a valuable new aid in the language classroom, one that can augment traditional learning if implemented thoughtfully, but it requires a rethinking of instructional approaches to ensure that the ultimate goal remains learners' autonomous communicative competence, not just error-free texts produced by a black box.

4.3 RQ3. What Implications Does This Evidence Offer for Future Research, Pedagogical Implementation, and the Evolving Role of AI in EFL Translation Learning?

Across the 26 included studies, a clear implication is that pedagogical value does not come from AI access alone, but from how tool use is structured, evaluated, and regulated in learning activities. Rather than framing AI translation tools as either "helpful" or "harmful," the evidence consistently points to conditional benefits: gains are more likely when learners engage critically with AI output, articulate reasons for revisions, and treat AI as a resource to be interrogated. This emphasis on guided engagement appears in both MT-focused studies and more recent LLM-oriented work, where learners' tool use is widespread but their ability to evaluate output quality varies substantially (Vukanović, 2025; Dinh, 2025; Klimova, 2025).

A first implication for pedagogical implementation is the need to formalize "AI translation literacy" as an instructional target. Several studies suggest that learners want explicit guidance on how to use translation tools effectively, which implies that leaving tool use to informal self-help may amplify convenience without ensuring learning (Yang et al., 2023; Daniel & Ancheta, 2025). In practice, this means embedding tool-use instruction into course routines, including when to consult AI, how to verify adequacy and appropriateness, and how to justify post-edits. Dinh (2025) illustrates this shift particularly clearly in LLM settings, where students' outcomes depend on how they interact with the tool, including the quality of prompts and the extent to which they treat AI output as a draft that requires verification and revision. Similarly, Daniel and Ancheta (2025) highlight that students often use Google Translate strategically as a lexical resource, which suggests a teachable pathway: instructors can promote selective, segment-level consultation while discouraging wholesale copying. At the curriculum level, Li and Tekwa (2025) further indicate that MT can be incorporated as a cognitive tool in translation and CAT-related instruction when teachers make strategy use explicit and provide feedback structures that support self-regulation, rather than treating MT use as a purely individual habit.

A second, closely related implication is that post-editing and error-focused comparison tasks are central mechanisms for turning AI output into learning. Multiple studies imply that translation pedagogy should reposition AI output as an object for critique, not a final product. In translator training contexts, Varela Salinas and Burbat (2023) argue for integrating MT and post-editing as a normal component of instruction and assessment, emphasizing systematic diagnosis of typical MT errors and targeted reinforcement. Similar instructional logic appears in proficiency-focused studies suggesting that post-editing tasks can stimulate deeper processing, especially when students are required to explain revisions rather than merely "fix" surface errors (Kim & Cha, 2023; Chung, 2020). In more recent comparative work, Abdelhalim et al. (2025) shows that students' evaluations of ChatGPT and Google Translate are shaped by perceived strengths and limitations, which supports a pedagogy that requires learners to compare tool outputs, identify where each tool performs well, and defend their choice of revisions using translation-relevant criteria. Chon et al. (2021) likewise indicate that while MT can support linguistic sophistication, it can also introduce mistranslations and poor lexical choices, implying that classrooms should integrate meaning-checking routines and guided editing so that learners learn to detect and correct these predictable weaknesses. From this perspective, the "productive" instructional use of AI is not simply to obtain better drafts, but to make learners practice evaluation, decision-making, and accountability, which aligns with the repeated observation that uncritical acceptance of AI output increases the risk of shallow processing (Garcia & Pena, 2011; Ali, 2022).

A third implication concerns differentiation by learner characteristics, particularly proficiency and evaluative capacity. Several studies suggest that the same tool can function as a scaffold for one learner group and as a dependency risk for another. Kim and Cha (2023) explicitly point to proficiency-sensitive implementation when using MT with post-editing, implying that weaker learners may require stronger scaffolding to avoid over-trusting output and to keep attention on meaning, while stronger learners may benefit from more complex editing demands and higher expectations for justification. In a similar vein, Chung (2020) suggests that the effectiveness of MT post-editing as instruction depends on proficiency and task design, reinforcing the implication that "one-size-fits-all" integration is unlikely to be optimal. Evidence from tool-output analyses also supports differentiation: Huang et al. (2025) indicate that LLM and MT outputs can differ in linguistic complexity, which implies that teachers may need to align tool choice and task demands with learners' developmental readiness. This proficiency-sensitive logic is also consistent with earlier work showing that beginners may experience increased fluency and participation through MT assistance but also worry about reduced independent thinking if tool use replaces their own linguistic effort (Garcia & Pena, 2011). Consequently, the pedagogical implication is not to prohibit tools for lower proficiency learners, but to design tasks that require active engagement, such as limited-scope tool use, mandatory post-editing steps, or reflective rationales that keep learning goals in focus.

A fourth implication is that responsible implementation requires explicit norms, institutional guidance, and integrity-oriented classroom policies. Several studies raise concerns about overreliance, inaccurate output, and misuse, suggesting that responsible-use frameworks should be part of teaching rather than an external administrative add-on. Nugroho et al. (2025) emphasize benefits but also highlight risks associated with credibility and uncritical uptake, implying the need for verification routines and clear expectations for attribution and acceptable use. Bekdaş (2025) similarly underscores that while learners may value ChatGPT's convenience, concerns about inaccuracy, dependency, and academic dishonesty remain salient, supporting explicit discussion of responsible use and strategies for checking reliability. In secondary contexts, Alghasab (2025) reports that learners may be uncertain about whether AI use strengthens or weakens independent ability over time, which implies that teachers should balance AI-assisted practice with tasks designed to maintain autonomy and self-generated production. In a broader adoption-oriented study, Wang and Wang (2025) suggest that tool effectiveness in low-resource settings depends not only on technical improvement (e.g., better translation quality) but also on learners' expectations and social influences, implying that institutional support and clear implementation guidance can shape both uptake and learning outcomes. Taken together, these studies imply that the evolving classroom role of AI requires governance through transparent policies and explicit pedagogical routines that make evaluation, accountability, and ethical considerations visible (Aijun, 2024; Wong & Symons, 2025).

In addition to implications for teaching, the evidence provides a coherent agenda for future research. A dominant implication is methodological: many studies rely on self-report, short-term outcomes, or context-specific samples, which limits claims about durable translation-skill development. Klimova (2025) and Pikhart et al. (2024) both point to the need for broader designs beyond snapshot surveys, implying that future work should incorporate longitudinal tracking, stronger comparison groups, and outcomes tied to translation competence rather than general attitudes alone. Process-oriented designs are also repeatedly implied as necessary. Chon et al. (2021) suggest that understanding how learners actually work with MT requires process evidence (e.g., examining the gap between tool output and learners' post-edited versions), and Dinh (2025) similarly implies that interaction patterns (including iterative prompting in LLM contexts) are central to explaining why some learners benefit more than others. Studies focusing on meaning-sensitive or culture-loaded content imply another research direction: Vukanović (2025) and Abdelhalim et al. (2025) both indicate that figurative language and literary translation demands reveal tool limitations that standard accuracy metrics may miss, suggesting future research should test a wider range of genres and culturally embedded texts with translation-relevant criteria and detailed error typologies. In parallel, newer work highlighting "translationese" or genre-insensitive patterns in AI outputs implies that future studies should develop stronger benchmarks for naturalness, pragmatics, and appropriateness, and examine how prompt conditions and interactive workflows shape these outcomes (Kwok et al., 2025; Alkhawaja, 2024).

A major methodological limitation across the 26 included studies is the predominance of short-term or cross-sectional designs. Most interventions ranged from single-session tasks to semester-based implementations, which restricts claims regarding sustained translation competence development. While AI-assisted conditions frequently yielded improved immediate performance, the absence of delayed post-tests and transfer measures limits conclusions about durable skill acquisition (Pikhart et al., 2025; Klimova, 2025). Future research should prioritize longitudinal designs incorporating delayed assessments, tool-free transfer tasks, and multi-semester tracking of learners' independent translation performance. Without such evidence, observed gains may reflect performance augmentation rather than developmental change. Establishing durable competence transfer therefore represents a critical next step for the field.

Finally, the literature converges on an evolving role of AI in EFL translation learning, shifting from a convenience tool toward a mediated learning partner and a catalyst for redefining translation competence. Across studies that discuss translator training and classroom integration, AI is not positioned as a replacement for learning or for human judgment; instead, it increases the importance of post-editing, evaluation, and workflow management as core competencies. Varela Salinas and Burbat (2023) explicitly connect MT integration to changes in training priorities, emphasizing post-editing and critical review as increasingly central professional and educational skills. LLM-focused studies similarly imply that translation competence now includes the ability to interact effectively with AI, to refine inputs, and to supervise outputs through verification and culturally sensitive revision (Dinh, 2025; Kwok et al., 2025; Alkhawaja, 2024). This reframing is also consistent with findings that learners can benefit from AI support, but only when instruction keeps learners accountable for meaning, appropriateness, and justification of choices (Daniel & Ancheta, 2025). Overall, the synthesized evidence implies that future EFL translation pedagogy should treat AI as a normal but disciplined component of learning, where the key outcomes are not only improved products, but strengthened evaluative judgment, post-editing competence, and self-regulated tool use that can transfer beyond any single platform (Yang et al., 2023; Li & Tekwa, 2025; Kim & Cha, 2023).

4.4 ETLS Based Synthesis Summary

Viewed through the ETLS framework, the reviewed evidence reveals consistent patterns across dimensions. In terms of Environment, most studies examined freely accessible AI tools, with neural machine translation systems (e.g., Google Translate) dominating earlier work and increasing attention to large language models (e.g., ChatGPT) in recent studies. Regarding Task, translation activities were typically short-text or paragraph-level, predominantly L1-L2 (EFL) translation, with limited genre diversification beyond general academic or instructional texts. The Learner dimension was characterized primarily by tertiary-level EFL students, with proficiency emerging as a key moderator of outcomes. Finally, the Strategy dimension showed the greatest convergence, with recurring reports of post-editing, cross-tool comparison, selective tool use, reverse translation checks, and iterative refinement, indicating that strategic engagement mediates the effectiveness of AI-assisted translation learning.

5. Discussion and Conclusion

This systematic review synthesized evidence from 26 empirical studies on how AI translation tools shape EFL learners' translation learning and performance. Using the ETLS framework to organize Environments, Tasks, Learners, and Strategies, the review provides an integrated account of what learners do with AI tools, what outcomes have been observed, and what these findings imply for research and pedagogy. Across the corpus, two broad conclusions emerge. First, AI-assisted translation has become a routine feature of EFL translation learning environments, dominated by free and easily accessible NMT tools (notably Google Translate) and increasingly complemented by LLM-based systems such as ChatGPT in the newest studies (Klimova, 2025; Dinh, 2025; Bekdaş, 2025). Second, the evidence supports a conditional effectiveness view: AI tools can enhance task performance, but learning value depends on the strategies learners employ and the degree of instructional structuring that converts tool use into reflective, competence-building activity (Chon et al., 2021; Kim & Cha, 2023; Li & Tekwa, 2025; Kruk & Kałużna, 2025).

With respect to RQ1, the review consolidates a coherent pattern of learner-reported AI-assisted translation strategies that extends beyond simple "use vs. non-use." Across MT- and LLM-oriented studies, learners commonly reported cross-checking outputs across tools and resources, iterative post-editing of AI drafts, selective use for difficult segments (e.g., lexical or phrase-level support rather than full-text substitution), and verification routines such as back-translation or meaning confirmation (Vukanović, 2025; Daniel & Ancheta, 2025; Garcia & Peña, 2011). Importantly, several studies imply that these strategies are not merely operational habits but reflect cognitive and metacognitive regulation, including monitoring tool reliability, managing dependence, and aligning tool use with task goals (Dinh, 2025; Yang et al., 2023). By synthesizing these recurring behaviors, the review contributes a clearer strategic vocabulary for describing AI-mediated translation learning, which can support more systematic process-oriented research and more explicit classroom strategy instruction.

Regarding RQ2, the evidence suggests that AI translation tools often improve immediate performance outcomes in translation-related tasks, though effects vary by task demands, learner proficiency, and evaluation criteria. Multiple studies reported higher quality products in AI-assisted conditions, such as fewer grammatical errors, more fluent output, and improved lexical choice, which is consistent across both earlier MT research and newer work on generative tools (Garcia & Peña, 2011; Alghasab, 2025). At the same time, the review clarifies why results sometimes appear mixed. Some studies indicate that while linguistic sophistication may rise, meaning accuracy can suffer through mistranslations, inappropriate word choice, or pragmatic mismatch, underscoring the need to separate "surface improvement" from "communicative adequacy" (Chon et al., 2021; Ali, 2022). Evidence also indicates that proficiency moderates outcomes: lower-proficiency learners may benefit from scaffolding that helps them produce more complete and accurate texts, yet they may also face higher risk of over-trust when they lack resources to evaluate AI output (Garcia & Peña, 2011; Kim & Cha, 2023). Taken together, the review's synthesis suggests that the most robust pattern is not uniformly positive or negative, but rather that AI tools enhance performance most reliably when learners engage in evaluative and revision-oriented strategies and when tasks require accountable editing rather than passive acceptance.

For RQ3, the synthesized evidence yields implications that are both pedagogical and theoretical. Pedagogically, the studies converge on the need to treat AI translation tools as disciplined learning resources embedded in instruction, not informal shortcuts used outside the pedagogical frame. Several studies highlight learners' demand for explicit guidance and indicate that unstructured use can amplify convenience without ensuring learning; therefore, "AI translation literacy" should become a curricular objective, including knowing when to use AI, how to verify output, and how to justify revision decisions (Yang et al., 2023; Nugroho et al., 2025; Daniel & Ancheta, 2025). A consistent instructional implication is to center post-editing and comparison tasks as mechanisms that transform

AI output into learning opportunities. This is emphasized in MT-oriented translator education where MT and post-editing are framed as legitimate competencies and error diagnosis becomes a teachable routine (Varela Salinas & Burbat, 2023; Chung, 2020). Similar logic is visible in LLM settings, where outcomes depend strongly on interaction practices, including prompt quality, iterative refinement, and verification, reinforcing the need for structured classroom workflows rather than one-off tool use (Dinh, 2025; Kwok et al., 2025).

The evidence further implies that implementation should be differentiated. Proficiency-sensitive designs are repeatedly recommended, because learners' ability to evaluate and revise AI output determines whether AI serves as scaffold or substitute (Kim & Cha, 2023; Chung, 2020). In parallel, many studies emphasize governance issues such as overreliance, inaccurate output, and academic integrity risks, implying that responsible-use norms and transparency expectations should be addressed explicitly in teaching and assessment rather than treated as peripheral concerns (Bekdaş, 2025; Nugroho et al., 2025; Wong & Symons, 2025). These concerns are especially salient as tool quality improves: the more fluent AI output becomes, the more learners may confuse plausibility with correctness, making verification and accountable authorship central learning outcomes.

Conceptually, the review suggests that AI is reshaping translation learning from a product-only orientation to a hybrid competence model where evaluation, post-editing, and workflow management become core skills. Translator training research explicitly frames this shift as a change in professional and educational profiles toward post-editing and technology-mediated language work (Varela Salinas & Burbat, 2023). More recent LLM-oriented studies similarly imply that learners must develop competencies for interacting with AI as a collaborative resource, while maintaining human oversight for pragmatics, culture, and task purpose (Kwok et al., 2025; Alkhawaja, 2024). Thus, the evolving role of AI in EFL translation learning is best understood not as replacing translation competence, but as increasing the importance of judgment, verification, and revision literacy.

While consolidating current knowledge, the review also identifies priorities for future research grounded in the included studies. First, the predominance of short-duration studies further underscores the need for longitudinal evidence to establish whether AI-assisted improvements translate into durable translation competence. Second, future work should adopt more process-sensitive methods to capture how learners interact with AI, including how they select segments for tool use, how they post-edit, and how verification strategies mediate outcomes (Chon et al., 2021; Dinh, 2025). Third, research should expand scope across genres and meaning-sensitive content, because several studies indicate that figurative language, cultural nuance, and pragmatic appropriateness remain recurring challenge areas where evaluation and human judgment are most visible (Vukanović, 2025; Ali, 2022). Finally, given the emerging role of LLMs and concerns about "AI translationese" and genre sensitivity, research should examine human–AI workflow designs and instructional prompting practices that shape output quality and learning value (Kwok et al., 2025; Alkhawaja, 2024).

6. Co-Author Contribution

The authors declare that there is no conflict of interest regarding this article. Author 1 conceptualized the study, developed the framework, designed the methodology, supervised the research process, and oversaw the overall manuscript preparation. Author 2 conducted the literature search, screening, data extraction, and prepared the initial draft of the manuscript under the supervision of Author 1. Both authors reviewed and approved the final version of the manuscript.

7. Declaration of AI-Generated Content

Portions of the manuscript were developed with the assistance of generative AI tools for language refinement and structural editing. All conceptual development, data screening, synthesis, interpretation, and final content decisions were carried out and verified by the authors.

8. Acknowledgements

The authors would like to acknowledge the researchers whose studies contributed to thissystematic literature review. We also appreciate the institutional access provided to databases, which supported the literature search process.

9. References

Abdelhalim, S. M., Alsahil, A. A., & Alsuhaibani, Z. A. (2025). Artificial intelligence tools and literary translation: a comparative investigation of ChatGPT and Google Translate from novice and advanced EFL student translators' perspectives. *Cogent Arts & Humanities*, 12(1), 2508031.

Aijun, Y. (2024). On the influence of artificial intelligence on foreign language learning and suggested learning strategies. *International Journal of Education and Humanities*, 4(2), 107-120.

Alghasab, M. B. (2025). English as a foreign language secondary school students' use of artificial intelligence tools for developing writing skills: Unveiling practices and perceptions. *Cogent Education, 12*(1), 2505304. https://doi.org/10.1080/2331186X.2025.2505304

Ali, M. A. (2022). Machine translation in Arabic-speaking context and students' perceptions of it. *Eurasian Journal of Applied Linguistics, 8*(1), 343–350. https://doi.org/10.32601/ejal.911439

Alkhawaja, L. (2024). Unveiling the new frontier: ChatGPT-3 powered translation for Arabic-English language pairs. *Theory and Practice in Language Studies, 14*(2), 347–357. https://doi.org/10.17507/tpls.1402.09

Bada, S. O., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education, 5*(6), 66–70. https://doi.org/10.9790/7388-05616670

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv*. https://arxiv.org/abs/1409.0473

Bekdaş, M. (2025). The Pros and Cons of ChatGPT in Foreign Language Teaching and Its Impact on Student Motivation.

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics*, *19*(2), 263-311.

Cancino, M., & Panes, J. (2021). The impact of Google Translate on L2 writing quality measures: Evidence from Chilean EFL high school learners. *System, 98*, 102464. https://doi.org/10.1016/j.system.2021.102464

Chen, S., & Lin, Y. (2025). A multidimensional comparison of ChatGPT, Google Translate, and DeepL in Chinese tourism texts translation: Fidelity, fluency, cultural sensitivity, and persuasiveness. *Frontiers in Artificial Intelligence, 8*, 1619489. https://doi.org/10.3389/frai.2025.1619489

Chomicz, M. (2025, January). Enhancing EFL writing skills for adult deaf and hard of hearing individuals. In *Frontiers in Education* (Vol. 9, Article 1504503). Frontiers Media SA. https://doi.org/10.3389/feduc.2024.1504503

Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity, and errors. *System, 96*, 102408. https://doi.org/10.1016/j.system.2020.102408

Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *Journal of Asia TEFL, 17*(1), 182–196. https://doi.org/10.18823/asiatefl.2020.17.1.12.182

Daniel, D., & Ancheta, R. (2025). A systematic review on the utilization of Google Translate in learning English as a foreign language. *European Journal of Applied Linguistics Studies, 8*(4).

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Deng, X., & Yu, Z. (2022). A systematic review of machine-translation-assisted language learning for sustainable education. *Sustainability, 14*(13), 7598. https://doi.org/10.3390/su14137598

Dinh, C. T. (2025). EFL Students' Perspectives on ChatGPT in Translation: Exploring AI Assistance, Motivation, and Learning Outcomes. *Electronic Journal of e-Learning*, *23*(2), 99-116.

Dostert, L. E. (1955). The georgetown-ibm experiment. *Machine translation of languages*, *8*, 129ff.

Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2001). Re-conceptualizing" scaffolding" and the zone of proximal development in the context of symmetrical collaborative learning. *The journal of classroom interaction*, 40-54.

García, I., & Pena, M. I. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning, 24*(5), 471–487. https://doi.org/10.1080/09588221.2011.582687

Holden, R. J., & Karsh, B.-T. (2009). A theoretical model of health information technology usage behaviour with implications for patient safety. *Behaviour & Information Technology, 28*(1), 21–38. https://doi.org/10.1080/01449290601138245

Huang, Y., Li, D., & Cheung, A. K. (2025). Evaluating the linguistic complexity of machine translation and LLMs for EFL/ESL applications. *Research Methods in Applied Linguistics, 4*(3), 100229. https://doi.org/10.1016/j.rmal.2025.100229

Hutchins, J. (2005). The history of machine translation in a nutshell. *Retrieved December*, *20*(2009), 1-1.

Jiang, L., Yu, R., & Zhao, Y. (2024). Theoretical perspectives and factors influencing machine translation use in L2 writing: A scoping review. *Journal of Second Language Writing, 64*, 101099. https://doi.org/10.1016/j.jslw.2024.101099

Kim, H. S., & Cha, Y. (2023). Post-editing of machine translation while reading on English proficiency levels. *Linguistic Research*, 40.

Kirchhoff, P. (2024). Machine translation in English language teaching. *ELT Journal, 78*(4), 393–400. https://doi.org/10.1093/elt/ccae026

Klimova, B. (2025). Use of machine translation in foreign language education. *Cogent Arts & Humanities, 12*(1), 2491183. https://doi.org/10.1080/23311983.2025.2491183

Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press.

Koka, N. A. (2024). The integration and utilization of artificial intelligence (AI) in supporting older/senior lecturers to adapt to the changing landscape in translation pedagogy. *Migration Letters*, *21*(S1), 59-71.

Kruk, M., & Kałużna, A. (2025). Investigating the role of AI tools in enhancing translation skills, emotional experiences, and motivation in L2 learning. *European Journal of Education, 60*(1), e12859. https://doi.org/10.1111/ejed.12859

Kwok, H. L., Shi, Y., Xu, H., Li, D., & Liu, K. (2025). GenAI as a translation assistant? A corpus-based study on lexical and syntactic complexity of GPT-post-edited learner translation. *System, 130*, 103618. https://doi.org/10.1016/j.system.2025.103618

Lee, S. M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning, 33*(3), 157–175. https://doi.org/10.1080/09588221.2018.1553186

Li, B., Lowell, V. L., Wang, C., & Li, X. (2024). A systematic review of the first year of publications on ChatGPT and language education. *Computers and Education: Artificial Intelligence, 7*, 100266. https://doi.org/10.1016/j.caeai.2024.100266

Li, C., & Tekwa, K. (2025). Translation as a catalyst for foreign language learning: a self-regulated learning approach mediated by instructor feedback and peer collaboration. *Humanities and Social Sciences Communications*, 12(1), 1-17. https://doi.org/10.1057/s41599-025-04766-3

McLeod, S. (2023). Constructivism learning theory & philosophy of education. *Simply Psychology*. https://www.simplypsychology.org/constructivism.html

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339*, b2535. https://doi.org/10.1136/bmj.b2535

Murtisari, E. T., Kristianto, A. K., & Bonar, G. (2024). Self-directed use of machine translation among language learners. *Foreign Language Annals, 57*(4), 1094–1114. https://doi.org/10.1111/flan.12768

Narayan, R., Rodriguez, C., Araujo, J., Shaqlaih, A., & Moss, G. (2013). Constructivism: Constructivist learning theory.

Nugroho, A., Andriyanti, E., Widodo, P., & Mutiaraningrum, I. (2025). Students' appraisals post-ChatGPT use. *Innovations in Education and Teaching International, 62*(2), 499–511. https://doi.org/10.1080/14703297.2024.2319184

O'Neill, E. M. (2019). Training students to use online translators and dictionaries. *International Journal of Research Studies in Language Learning, 8*(2), 47–65.

Pikhart, M., Klimova, B., & Al-Obaydi, L. H. (2024, July). Exploring university students' preferences and satisfaction in utilizing digital tools for foreign language learning. In *Frontiers in Education* (Vol. 9). https://doi.org/10.3389/feduc.2024.1412377

Reifler, E., & Pol, R. (1962). Machine translation. *Bulletin of the Medical Library Association, 50*(3), 473–480.

Salloum, S., Shaalan, K., Alfaisal, R., Salloum, A., & Gaber, T. (2024). Integrating ChatGPT into medical education: A combined SEM-ML approach. In *2024 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC)* (pp. 1–5). IEEE. https://doi.org/10.1109/ASSIC60049.2024.10507994

Shabani, K., Khatib, M., & Ebadi, S. (2010). Vygotsky's zone of proximal development: Instructional implications and teachers' professional development. *English Language Teaching, 3*(4), 237–248. https://doi.org/10.5539/elt.v3n4p237

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems, 27*.

Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence, 7*. https://doi.org/10.1016/j.caeai.2024.100270

Varela Salinas, M. J., & Burbat, R. (2023). Google Translate and DeepL: Breaking taboos in translator training.

Vukanović, M. B. (2025). Translating metaphors in the age of AI. *ELOPE: English Language Overseas Perspectives and Enquiries, 22*(1), 185–201. https://doi.org/10.4312/elope.22.1.185-201

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Wang, H., & Wang, X. (2025). Enhancing translation accuracy and learning outcomes for low-resource languages. *System, 134*, 103807. https://doi.org/10.1016/j.system.2025.103807

Wertsch, J. V. (2017). Mediated action. In *A companion to cognitive science* (pp. 518–525). Wiley.

Wong, E., & Symons, K. (2025). The Australian injectable drugs handbook: Experience of artificial intelligence translations of non-English product information. *Journal of Pharmacy Practice and Research, 55*(1), 68–78. https://doi.org/10.1002/jppr.1946

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv*. https://arxiv.org/abs/1609.08144

Yang, Y., Wei, X., Li, P., & Zhai, X. (2023). Assessing the effectiveness of machine translation in the Chinese EFL writing context. *ReCALL, 35*(2), 211–224. https://doi.org/10.1017/S0958344022000176

Zhang, J., Liao, Q., Li, L., & Luo, J. (2025). The influence of natural language processing on EFL speaking skills. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331251377414