ISSN 1798-4769 (print); ISSN 2053-0684 (online)
Journal of Language Teaching and Research, Vol. 17, No. 2, pp. 653-664, March 2026
DOI: https://doi.org/10.17507/jltr.1702.28

Exploring the Role of Large Language Models in Translation Education: A Systematic Review

Anas M. Alkhofi
Department of English, College of Arts, King Faisal University, Al-Ahsa 31982, Saudi Arabia

***Abstract*—Despite the surge of research interest in generative AI and the rapid public adoption of large language models (LLMs), their role in translation remains unclear. The reliability of these systems and their limitations as machine translation tools continue to be a central concern for translation teachers and students. Systematic reviews that specifically examine LLMs in translation are still scarce. This systematic review aims to address this gap by synthesizing and interpreting recent empirical studies on the use of LLMs in translation across three areas: (1) LLMs' translation quality, (2) LLM-generated translation feedback, and (3) the integration of LLMs into translation education. Drawing on 55 empirical studies, the findings show that LLMs—particularly GPT—consistently outperform conventional neural MT systems. For general, non-specialized texts, their output often approaches human quality, though human translators maintain a clear advantage in culturally dense, technical, or literary content. Evidence further indicates that LLMs can provide helpful and timely feedback that identifies common linguistic issues, which in turn can assist both teachers and students; however, teacher feedback remains superior in depth, contextual sensitivity, and clarity. As contemporary translation workplaces increasingly rely on MT and AI-supported tools, training students to work with LLMs has become essential for aligning classroom practice with professional expectations. At the same time, educators must balance LLM-assisted tasks with hands-on human translation to ensure that students continue to develop essential linguistic and problem-solving skills.**

***Index Terms*—ChatGPT for translators, LLMs in translation education, machine translation, translation feedback, translation assessment**

I. Introduction

While machine translation (MT) has been part of translation studies for decades, large language models (LLMs) such as ChatGPT represent a new generation of technology that has been adopted at an unprecedented pace across domains, including text generation, education, programming, and healthcare. This rapid uptake has been matched by a sharp rise in research interest: Zhang et al. (2025) report an annual growth rate of 545.53% in ChatGPT-related educational research, and Ng and Ho (2025) identify 3,808 peer-reviewed journal articles on generative AI in education between 2022 and 2025, describing this trend as "exponential growth".

Because this technology is still relatively new—first made publicly available with OpenAI's ChatGPT model in late 2022 (OpenAI, 2022)—there remains considerable uncertainty about what these tools can actually do and how they should be used in translation education. Alghamdi and Alotaibi (2025) noted that a clear gap remains in the literature regarding the pedagogical use of generative AI (GenAI) tools in translation. The pace of technological development appears to have surpassed educators' understanding of how LLMs should be integrated into translation education and training. Teachers and students alike are still learning how reliable these systems are and what their limitations might be. Researchers have emphasized that GPT's pedagogical potential is still emerging (Mizumoto & Eguchi, 2023; Gjorevski et al., 2025) and that drawing overarching conclusions from previous studies remains challenging for educators (Lee, 2023). Accordingly, the goal of this systematic review is to synthesize and interpret recent empirical studies on the use of LLMs in translation. Specifically, it addresses three key questions: (1) How do LLMs perform in terms of translation quality? (2) To what extent do LLMs provide useful feedback on translations? and (3) What does existing research reveal about incorporating LLMs into translation pedagogy?

II. Literature Review

A. Stages of Machine Translation

Machine translation research began in the late 1940s, when Weaver (1947) proposed using computers to translate languages (Weaver, 1952). Early attempts led to rule-based MT (RBMT, 1950s–1990s), where systems relied on dictionaries, morphological analyzers, and thousands of hand-crafted rules. These systems required enormous manual effort, and the rules were difficult to scale across domains or languages. At first, they were mainly designed for military applications. Later, as the systems matured, commercial use began in 1978 with SYSTRAN, and even as late as 2007, Google adopted rule-based MT (Wang et al., 2022).

In the 1990s, attention shifted toward data. Statistical MT (SMT) was proposed by Brown et al. (1990). SMT relied on probabilities from parallel texts. Unlike RBMT, it automatically learned translation knowledge from large amounts of

data instead of depending on human experts to hand-write rules. SMT methods became widely adopted as they matured. In 2006, Google launched its internet translation service based on SMT, and in the years that followed, other companies such as Microsoft and Baidu also introduced similar services. However, SMT often failed to handle languages with very different word orders, which led to awkward or incorrect structures in translations between distant language pairs (Wang et al., 2022).

With the rise of neural networks, MT entered a new phase. The work of Bahdanau et al. (2014), Sutskever et al. (2014), and Dong et al. (2015) introduced neural machine translation (NMT), and large-scale deployments soon followed, with Google releasing its neural systems in 2016 (Wu et al., 2016). NMT reads the entire source sentence, encodes it into a numerical representation, and then produces the translation word by word, more closely resembling how human translators first grasp meaning before rewriting it in another language. Trained on large bilingual corpora rather than hand-written rules or word-level probabilities, NMT generally produces more fluent and natural output (Ataman et al., 2025). It has therefore become the dominant MT approach and triggered debates about whether MT quality can be considered comparable to human translation.

Most recently, the emergence of LLMs has brought another shift in MT. While LLM research began earlier, it was the public release of OpenAI's models in 2022 that marked the true beginning of LLMs on a mass scale (OpenAI, 2022). Unlike earlier neural systems that were trained specifically for translation, LLMs such as GPT-4 or Gemini are trained on massive multilingual and multi-domain datasets and can perform translation as part of their broader text generation abilities. These models show strengths in producing fluent and context-aware translations, sometimes handling discourse and style better than traditional NMT. Because of their flexibility and integration into many applications, LLMs now represent the latest stage in MT evolution.

B. Previous Systematic Reviews on LLMs in Education

Although research on LLMs in translation has been growing rapidly, systematic reviews that synthesize this emerging body of work remain limited. To date, the only systematic review focusing specifically on LLMs in translation is that of Chan and Tang (2024). They conducted a systematic review of 26 studies to identify the main research trends, themes, and outcomes related to GPT in translation. Their analysis revealed a substantial rise in publications. The findings showed that GPT-produced translations often match human translations in quality and outperform conventional neural MT outputs. GPT also demonstrated strength in handling complex linguistic features, such as humor, puns, poetry, and cultural references. Moreover, the review highlighted GPT's potential for tasks beyond translation itself, including post-editing and translation quality evaluation. Importantly, the study emphasized that GPT's performance is highly dependent on prompt design, with detailed and context-specific prompts (e.g., including information on the target audience, translation purpose, or examples) leading to higher accuracy.

While systematic reviews on LLMs and translation are limited, several systematic reviews have focused on MT more broadly. For instance, Tafa et al. (2025) systematically analyzed 69 articles to assess MT performance in low-resource languages. Their review found that MT performance remains constrained by data scarcity and the structural complexity of many low-resource languages. However, they noted that the emergence of large language models, including GPT-4, is reshaping the MT landscape through zero-shot and few-shot learning capabilities, which allow for cross-lingual generalization even without parallel corpora.

Similarly, Almaaytah and Alzobidy (2023) addressed the linguistic and technical difficulties involved in rendering Arabic into English using MT. Their systematic review revealed that Arabic–English translation poses unique challenges such as word sense disambiguation, named entity recognition, complex morphology, and low-resource data availability. Word sense disambiguation becomes problematic when missing diacritics lead MT systems to misinterpret words with multiple meanings, producing inaccurate or awkward translations. Likewise, the absence of capitalization in Arabic makes it difficult for MT to recognize proper names, reducing consistency in named entity translation. The language's rich and complex morphology, where single words can encode multiple grammatical elements, further confuses segmentation and alignment processes. Finally, as Arabic remains a low-resource language, limited parallel training data restricts model learning, resulting in less accurate and less natural outputs. The study highlighted that these intertwined linguistic and resource challenges continue to hinder MT performance for Arabic–English translation.

In another systematic review, Rivera-Trigueros (2022) examined 27 studies to determine which MT systems are most commonly used, their underlying architectures, and the quality assessment methods applied. The results indicated that neural MT, especially Google Translate, dominates the current research landscape. Most studies relied on either automatic or human evaluation alone, while only a small portion (22%) combined both methods. Over half of the studies incorporated detailed error analysis, which Rivera-Trigueros identified as crucial for improving system performance and understanding translation limitations.

Despite growing academic interest and widespread adoption of GPT and other LLMs in translation practice (Chan & Tang, 2024), systematic reviews summarizing their use and pedagogical value remain limited. Therefore, the present study aims to address this gap by reviewing recent empirical research on three interconnected themes: (1) LLMs and translation quality, (2) LLMs and translation feedback, and (3) LLMs and translation education. This synthesis seeks to provide a clearer understanding of current research directions and inform future pedagogical and technological developments in translation education.

III. Methodology

This study employed a systematic review methodology to examine recent empirical research on LLMs and translation. This review followed the well-established PRISMA guidelines for reporting systematic reviews (Page et al., 2021). The review covered studies published from 2016 onward. However, since the widespread public availability of LLMs began in 2022, the majority of the reviewed studies were published between 2022 and 2025. The investigation was guided by three research questions:

RQ1. How do LLMs perform in terms of translation quality?

RQ2. To what extent do LLMs provide useful feedback on translations?

RQ3. What does existing research reveal about incorporating LLMs into translation pedagogy?

The procedure for identifying, selecting, and screening relevant studies was carried out as follows.

A. Research Focus

This review was guided by the practical and pedagogical concerns of translators and translation educators, aiming to synthesize empirical evidence on three central themes: (1) LLMs and translation quality; (2) LLMs and translation feedback; and (3) LLMs and translation education.

B. Search Strategy and Databases

The search process was conducted across three major databases: Google Scholar, Scopus, and Web of Science. The search string required that article titles contain at least one term referring to LLMs or machine translation and one term referring to translation. Specifically, titles had to include: Large language model OR LLM OR GPT OR ChatGPT OR machine translation AND translate, translation, translated, or translating.

Across the three databases, this initial search generated approximately 4,000 records (see Figure 1). To ensure relevance to translation pedagogy and applied linguistics rather than computer science or engineering, results were restricted to peer-reviewed journal articles, English-language publications, and subject areas within the social sciences, linguistics, and humanities. After applying these filters, we identified 517 articles in Web of Science, 462 in Scopus, and 108 in Google Scholar. A total of approximately 1087 records were carried forward to the screening stage.

C. Screening and Selection Process

The titles and abstracts of retrieved studies were manually screened to identify those aligning with the review's objectives. Articles meeting the inclusion criteria were listed in a Word document table that captured the title, authors, assigned theme, and main findings.

Studies were included if they:

1. Had a title clearly addressing large language models or machine translation in relation to translation.
2. Explicitly aimed to empirically evaluate the translation performance of LLMs—either in comparison to human or neural MT—or reported specific shortcomings of LLM outputs.
3. Empirically examined the quality or pedagogical effectiveness of LLM-generated translation feedback from the perspective of students or teachers.
4. Investigated the application or pedagogical use of LLMs or MT within translation classrooms.
5. Were empirical primary research studies. Conceptual papers, opinion pieces, and literature reviews were excluded, as were computer science–oriented technical studies focused on model engineering or system optimization.

D. Data Extraction and Refinement

Screening and adding papers to the Word document table continued iteratively across databases until thematic saturation was reached—when additional searches and pages yielded no new relevant studies. At this point, 55 articles were identified; the remaining records either did not fit the themes or were duplicates appearing in two or all three databases. Of the 55 included articles, 34 addressed translation quality (first theme), 4 examined translation feedback (second theme), and 17 focused on translation education (third theme).

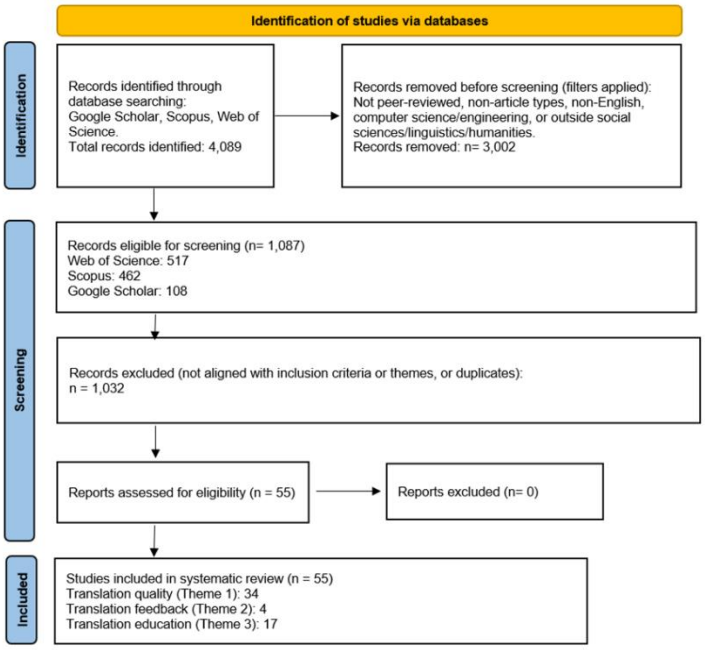

Figure 1. PRISMA Flow Diagram of the Study Selection Process

IV. Results

The purpose of this systematic review was to synthesize current empirical research on the emerging role of LLMs in translation. The following section presents the main findings and general trends identified across the reviewed studies, organized according to the three thematic areas: (1) LLMs and translation quality, (2) LLMs and translation feedback, and (3) LLMs and translation education.

A. LLMs and Translation Quality

(a). LLMs Compared to Humans

When human translation is used as the reference point, current evidence shows that LLMs still lag behind, though not by wide margins. In fact, some studies report a competitive performance by GPT compared to human translators, particularly in terms of fluency and surface quality (Woodrum, 2024; Haider & Alkhatib, 2024; Calvo-Ferrer, 2024). At the same time, adequacy and cultural sensitivity remain areas where human expertise holds an advantage.

Several studies highlight cases where GPT's output rivals that of human translators. For example, Woodrum (2024) reports that short English passages rendered into Mandarin by GPT were of statistically comparable quality to professional human translations, and Toledo-Báez (2024) also found that the type of translation (human or MT) did not significantly affect perceived quality. Similarly, Haider and Alkhatib (2024) conclude that GPT performs almost as well as humans in the translation of English legal acronyms into Arabic, a domain where terms usually have clear, one-to-one matches between languages. Calvo-Ferrer (2024) further shows GPT's strength in tasks aimed at general readers: translation students could not reliably distinguish between GPT-generated and human-generated subtitles, even when humor, irony, and cultural references were involved. Similarly, Alkhofi (2025) shows that even university professors could not reliably tell apart MT from student translations, and in many cases, they even preferred the MT output. Together, these findings underscore GPT's potential in tasks where fluency and surface naturalness are central.

At the same time, human translators continue to hold an advantage in contexts requiring literary creativity and precise meaning. Farghal and Haider (2024) report that, although GPT approximates human performance in translating Classical Arabic verse by capturing rhyme and thematic clarity, humans still lead in overall creativity and prosody. Similarly, Al Rousan et al. (2025) show that human translation of Arabic literature attains higher adequacy (94.5%) than GPT (77.9%), even though GPT slightly surpasses humans in fluency (97.2% vs. 96.6%). Other studies confirm this pattern across genres and domains: Jiang (2025) highlights humans' superior ability to refine complex syntax; Moneus and Sahari (2024) emphasize more reliable handling of meaning, tone, and cultural nuance, which is crucial for domains like legal texts; and Lu et al. (2025) find GPT-4's performance comparable to professional translators across several languages but still caution against full substitution, recommending instead that LLMs be used as supportive tools within hybrid human–AI workflows (see also Manapbayeva et al., 2024).

Taken together, these studies reveal a pattern of context-dependent parity. GPT achieves results close to or indistinguishable from human translation in simple, general, or audience-focused domains, while human translators remain more reliable in domains demanding cultural nuance, semantic accuracy, and creative expression.

(b). LLMs Compared to NMT

LLMs such as ChatGPT and NMT systems such as Google Translate have been evaluated across many contexts, and the evidence shows that LLMs are generally stronger. In most studies, LLMs produce translations that are more fluent, natural, and semantically faithful, while NMT systems often rely on literal renderings and word-level matches. Human judgments frequently prefer LLM outputs, especially for their ability to capture meaning and flow, though high-stakes uses still require human oversight.

In medical instructional texts, ChatGPT made fewer errors than Google Translate in Spanish (3.8% vs. 18.1%) and in Russian (35.6% vs. 41.6%). However, in Vietnamese, Google performed better (ChatGPT 24.2% vs. Google 10.6%). This shows that LLM advantages do not apply equally across all contexts (Rao et al., 2024). In technical translation, human evaluators preferred ChatGPT-4 over Google, even though automatic metrics such as BLEU favored NMT (Zhang et al., 2025; Jiang & Zhang, 2024). A broader comparison of different MT systems placed ChatGPT-4 ahead of Google, Bing, Yandex, Systran, and Amazon when judged against human references (Habib et al., 2025). Directionality also matters: all systems perform better from English into Chinese than the reverse, and ChatGPT often yields smoother and more readable Chinese to English translations (Cai, 2024).

When translating public speeches and literary material, LLMs usually preserve consistency, imagery, and flow more effectively. Google's Arabic to English translations of speeches required major revisions, while ChatGPT outputs were judged acceptable with minor edits (Alafnan, 2025). Student raters likewise valued ChatGPT's literary translations as richer and more consistent than Google's, even though some cultural references and figurative devices remained difficult for the model (Abdelhalim et al., 2025). Earlier studies also observed Google Translate's struggles with aesthetic and culture-bound expressions (Constantine, 2020; Al-Khresheh & Almaaytah, 2018). For scientific texts, ChatGPT offered clearer and more contextually accurate translations in comparison with Google Translate (Sadiq, 2025). Specialized domains confirm this advantage. For instance, in wine and olive oil tasting notes, ChatGPT-3.5 produced fewer terminological errors than Google Translate (Valdivieso & Arroyo, 2023). Research on Arabic poetic and culturally rich content further supports that LLMs usually deliver more fluent and culturally sensitive translations requiring fewer edits (Farghal & Haider, 2024).

Evaluation methods also shape the results. NMT systems often score higher on word-matching metrics such as BLEU, ChrF, and METEOR, which reward overlap at the word level. In contrast, LLMs perform better on semantic-oriented measures such as COMET and BERTScore. These metrics consider semantic similarity and fluency rather than focusing only on exact word matches. Human evaluators also tend to prefer LLM outputs for these reasons (Jiang & Zhang, 2024; Zhang et al., 2025). Indeed, Huang et al. (2025) show that LLMs often produce translations with greater lexical and syntactic complexity, offering richer structures than the simpler translations generated by NMT systems.

Overall, the evidence suggests that LLMs usually surpass NMT in fluency, readability, and semantic accuracy. NMT systems, however, still show strengths in areas that demand exact wording, terminological consistency, and alignment with reference corpora.

(c). Areas Where LLMs Still Struggle

Although LLMs increasingly approach human-level performance and often surpass neural machine translation systems, they still face notable challenges. These appear mainly in domains that require cultural awareness, specialized knowledge, or creative adaptation. Studies repeatedly point to weaknesses in literary translation, culturally dense material, low-resource languages, and specialized discourse.

One theme across studies is the persistent struggle of LLMs with literary and creative texts. Research shows that LLMs encounter significant difficulties in poetry and drama, where pragmatic, rhetorical, and cultural sensitivity are essential (Ed-Dali, 2025). In Persian-to-English translation of short stories, LLMs show problems with cultural nuance and idiomatic expressions, achieving only moderate accuracy even when outperforming NMT (Aghai, 2024). In Arabic literary work, accuracy remains well below that of human translators, averaging 77.9% for ChatGPT compared to 94.5% for human translation, with issues such as unnecessary additions and failures in handling structural elements (Al Rousan et al., 2025). Studies of fantasy novels confirm that manipulated idioms remain a challenge, where human creativity is far superior (Corpas Pastor & Noriega-Santiáñez, 2024). Religious translation also exposes weaknesses, with LLMs failing to capture nuance, introducing redundancy, and producing cases of Englishization in the renderings (Shormani & Alfahad, 2025). These findings confirm that LLMs, while fluent, lack the depth needed for literary and religious translation.

A second recurring weakness relates to cultural and idiomatic translation. Evidence shows that LLMs struggle with Chinese-to-English translations of abbreviations and idioms, though they can handle well-known set expressions more effectively (Wu, 2023). Failures in rendering cultural aspects accurately are also documented (El-Saadany, 2024). In specialized registers such as law, medicine, and literature, reliability remains low (Mohsan & Nayab, 2024). Collectively, these results show that cultural nuance and domain-specific knowledge continue to present significant obstacles.

LLMs also show shortcomings in scientific and technical texts. Both LLMs and NMT still require substantial training to manage scientific discourse effectively (Alzain et al., 2024). Significant variation across systems emphasizes the need for human translators in editing and verification (Habib et al., 2025). Specialized domains such as medicine and law continue to highlight these weaknesses (Mohsan & Nayab, 2024).

Another important limitation concerns language coverage, directionality, and sensitivity to context. LLMs, like other MT systems, perform better in high-resource European languages than in low-resource or typologically distant languages (Jiao et al., 2023; Zhang, 2022), and they generally achieve higher quality when translating from English into another language than in the reverse direction (Cai, 2024). Additional challenges arise in structurally complex languages such as Arabic and Persian (Ed-Dali, 2025; Aghai, 2024). Translation quality is also highly dependent on prompting: studies show that detailed, context-rich prompts improve accuracy and cultural appropriateness, particularly in literary, media, and Chinese–English tasks (Sadiq, 2025; Gao et al., 2024). These findings indicate that LLM reliability is shaped by language resources, translation direction, and the quality of contextual input provided.

Overall, the evidence shows that LLMs, despite their progress, are not yet reliable replacements for human translators in areas where cultural nuance, creativity, religious sensitivity, or technical expertise are required. Their outputs often need revision, and their performance varies depending on language resources and the quality of context provided.

B. LLMs and Translation Feedback

Research on LLMs as feedback tools in translation education remains limited, yet a growing body of work has begun to explore their potential (Su et al., 2025; Xu et al., 2025). Evidence indicates that ChatGPT can generate a considerable number of evaluative comments; however, the overall quality and reliability of this feedback vary, and studies caution against assuming parity with teacher input at this stage.

Su et al. (2025) compared ChatGPT and teacher feedback on Chinese-to-English translation tasks completed by master's students. They found that ChatGPT produced more comments than teachers, focusing mainly on lexical and grammatical issues and offering direct translations and general corrections. Teachers, on the other hand, provided fewer but more detailed and indirect comments that covered a wider range of translation aspects. Students viewed teacher feedback as clearer, more actionable, and more supportive of improvement. These findings suggest that while both feedback types can complement each other, too much general LLM feedback may overwhelm learners.

Research on learner engagement offers additional insight, showing that students interact actively but unevenly with LLM feedback. Surface-level comments on word choice or grammar are easily understood and acted upon, while higher-order advice on cohesion or coherence is harder to interpret when feedback remains too general. Students initially respond positively to ChatGPT's encouraging tone and generous scoring, which can enhance motivation, but over time, they prefer more critical, detailed guidance that facilitates deeper revision. Accordingly, teacher input continues to play a crucial role in addressing complex translation challenges (Xu et al., 2025).

Similarly, Cao and Zhou (2025) used BLEU scores to evaluate the effects of different feedback sources on students' translation revisions. When master's students revised Chinese–English translations based on self-, teacher-, or ChatGPT-generated feedback, the highest BLEU scores appeared in the self- and teacher-feedback conditions. Even so, ChatGPT feedback achieved higher scores than self and teacher feedback in terms of linguistic features, particularly in enhancing lexical capability and referential cohesion, whereas teacher feedback contributed more to sentence-level refinement.

Shifting toward a more encouraging perspective, Jiao et al. (2025) compared GPT-4 feedback with expert evaluations using English–Chinese student translations. Their findings revealed that LLM-generated feedback can identify a large proportion of the same issues noted by expert reviewers, thereby helping to alleviate the persistent feedback bottleneck created by teachers' limited time. In this study, 81.27% of ChatGPT's comments overlapped significantly with at least one expert's feedback, indicating that LLMs can approximate expert-like evaluation on a large scale in a structured format that supports instruction. Moreover, students expressed positive attitudes toward the LLM feedback, appreciating its flexibility, promptness, and accuracy in providing timely and useful comments.

Overall, the current evidence suggests that LLM-generated feedback can serve as a valuable supplementary resource for identifying common linguistic issues and providing timely support, while teacher feedback continues to add interpretive depth, contextual awareness, and task-specific direction (Su et al., 2025; Jiao et al., 2025; Xu et al., 2025; Cao & Zhou, 2025).

C. LLMs and Translation Education

Recent work increasingly indicates that integrating machine translation (MT) into translation education can be effective and pedagogically meaningful. A 16-week quasi-experiment comparing traditional instruction with MT-supported teaching reported statistically significant gains for the MT group on post-tests in both English-to-Chinese and Chinese-to-English translation, which indicated that neural machine translation tools substantially improved students' translation quality (Duan et al., 2025). In addition, Dinh (2025) showed that rather than spending time on routine translation tasks, GPT allows students to focus more on content comprehension and higher-order translation skills. Earlier classroom studies likewise reported that post-editing MT output can yield fluency and accuracy comparable to translating from scratch. They also found generally positive student attitudes toward post-editing (Jia et al., 2019).

Building on these outcomes, several authors argue for curricular redesign that cultivates students' MT competencies to mirror contemporary professional translation workflows. For example, Duan et al. (2025) recommended re-evaluating curriculum design and embedding NMT tools to develop students' AI competences; Wang (2023) argued that programs should reflect rapid technological change and prepare students for tool-rich professional practice; Krüger (2023) highlighted the need to build AI literacy among translators and terminologists; Johnston et al. (2024) called for universities to support students in learning to use AI productively and effectively; and Özmat and Akkoyunlu (2024) urged training

for students and faculty to ensure effective and ethical use of AI translation tools. Stapleton and Kin (2019) clarified that this parallel shifts in other disciplines. For example, in statistics, calculations once performed manually are now handled by software such as SPSS. In effect, this implies that MT might handle initial drafts, while human translators focus on refining, adapting, and quality assurance.

The case for the inclusion of MT tools within translator training is also grounded in current professional practice. Wang (2023) observed that professional translation workflows now rely heavily on CAT tools and MT services throughout projects, from communicating with clients to managing terminology and drafting. Koponen (2016) and Vieira (2019) noted that post-editing has become an established part of professional workflows rather than a niche activity, meaning that new translators will regularly encounter MT in their work. Likewise, Tavares et al. (2023) suggested that professional roles might evolve so that human translators increasingly handle complex or specialized tasks, while routine or repetitive segments are delegated to AI-supported systems. In line with these changes, Rico and Gonzalez Pastor (2022) recommended aligning coursework with this reality by incorporating MT and post-editing into assignments and assessments.

However, adoption of MT in education remains inconsistent. Although MT tools are increasingly common, appropriate pedagogical training for their integration is still rarely offered, leaving many instructors to improvise or avoid the technology altogether (He, 2021; Liu, 2018). This challenge is compounded by low educator MT literacy: many instructors report no formal preparation in MT or related tools, and few devote class time to demonstrating effective and ethical use (Rico & Gonzalez Pastor, 2022; Alkhofi, 2024). As several authors caution, easy access to technology does not automatically lead to competent use without explicit instruction and clear pedagogical guidance (Bowker, 2019).

A balanced view also points to risks and trade-offs associated with introducing MT tools in the classroom, especially at earlier stages of students' proficiency. Concerns include reduced diligence, weaker memory, and potential reduction of creativity when students over-rely on MT (Özmat & Akkoyunlu, 2024). Other researchers have echoed these concerns. For instance, Stapleton and Kin (2019) cautioned that heavy dependence on MT could reduce students' motivation to develop their own writing skills in the target language. In a similar vein, Lee (2020) observed that while MT can aid certain aspects of language learning, it may foster only a surface-level grasp of linguistic structure if not carefully administered by instructors. Scholars recommend maintaining a careful balance between AI-supported practices and hands-on translation to preserve active learning and professional judgement (Dinh, 2025). In practice, MT can be introduced alongside traditional hands-on activities, where some in-class tasks emphasize translation from scratch to build foundational skills, while others focus on post-editing MT outputs to strengthen analytical and revision abilities.

In translation education, post-editing is increasingly recognized as an essential practical skill that students can learn and apply. Surveys show that many professional translators now rely on MT post-editing in their daily work, highlighting its growing relevance in translator training (Povilaitienė & Kasperė, 2022). Moreover, experimental studies also demonstrate that post-edited MT translations can match or even surpass the quality of fully human translations while helping learners reduce frustration, time, and cognitive load (Yang et al., 2023).

That being said, post-editing remains a demanding task—it requires careful judgment and strong problem-solving skills, as students must learn to identify and correct subtle errors while managing the technical challenges of revision (Yao et al., 2025). In particular, because current neural and LLM-based MT often produce fluent-sounding writing, novice translators may overlook hidden mistakes, which underlines the importance of training that develops error awareness and evaluation skills (Yamada, 2019). Rico Pérez (2024) took this discussion further by calling for an update to existing post-editing guidelines, which were originally designed for older MT systems and are no longer suitable for today's more fluent models, which produce human-like errors. The proposed revisions aim to move beyond simple error correction toward a more dynamic approach that manages translation quality according to context, purpose, and expected outcomes, emphasizing active collaboration between human translators and MT systems.

V. Discussion and Conclusion

The purpose of this systematic review was to synthesize and interpret recent empirical evidence on the role of LLMs in translation practice and education. Guided by three research questions, it examined: (1) How do LLMs perform in terms of translation quality? (2) To what extent do LLMs provide useful feedback on translations, and (3) What does existing research reveal about incorporating LLMs into translation pedagogy? These questions were selected for their pedagogical and practical relevance, as they address issues that are increasingly central to translation educators seeking to understand and navigate the growing influence of artificial intelligence in their field.

With respect to the first research question, the evidence reviewed in this study shows that LLMs such as GPT demonstrate strong translation capabilities. When measured against professional human translators, LLMs generally lag behind, particularly in areas such as idiomatic accuracy, cultural sensitivity, and literary creativity. Human translators remain superior in rendering culturally specific meanings, figurative language, and stylistically nuanced texts. However, for general or non-specialized content, GPT's output often approaches human quality and is sometimes indistinguishable from professional translations.

When compared with NMT systems such as Google Translate, LLMs consistently outperform them. Studies report that GPT-generated translations are typically more fluent, cohesive, and natural in tone. Despite this improvement, several persistent challenges remain. LLMs still struggle with culturally bound expressions, specialized technical terminology,

and creative or metaphorical discourse. These weaknesses are particularly visible in literary or scientific domains that require advanced contextual interpretation. Another recurring limitation concerns directionality: LLMs perform better when translating from English into other languages than when translating into English, especially from low-resource languages. This asymmetry reflects the imbalance in available training data and linguistic representation. Overall, while LLMs have not yet reached full parity with human translators, they can function as powerful assistants—providing high-quality drafts that human translators can refine and enhance for final publication.

With regard to the second research question, findings indicate that LLMs can generate translation feedback that is largely acceptable and pedagogically useful, particularly for low-stakes or formative tasks. GPT-based systems can identify many of the same issues flagged by human evaluators, such as grammatical errors, lexical inaccuracies, and structural inconsistencies. However, human feedback continues to surpass AI-generated comments in depth, contextual sensitivity, and pedagogical value. Teachers tend to offer targeted explanations that guide learners toward understanding underlying translation principles, while LLM feedback often remains general and corrective rather than interpretive.

As such, GPT can serve as an effective supplementary tool for identifying common linguistic issues and encouraging self-revision, but it should not be considered a replacement for expert feedback. Human evaluators provide the interpretive depth and contextual awareness that allow for nuanced improvement in translation competence. LLMs, by contrast, are best positioned as scaffolding tools that support students in early drafting, reflection, or low-stakes assessment, where immediate, surface-level feedback is sufficient to foster revision and engagement.

In relation to the third research question, research consistently suggests that incorporating LLMs into translation teaching can be beneficial when done with pedagogical guidance. Experimental studies show that students' translation quality improves when MT or LLM tools are integrated into coursework. This integration aligns with professional practices, as machine translation and post-editing have become standard in the modern translation industry. Therefore, translation curricula should reflect this reality by preparing students to use such technologies effectively and critically.

Nevertheless, a key issue is that both teachers and students often lack formal training in how to use these tools productively and ethically. Without structured guidance, learners risk overreliance on automated systems, which may reduce creativity, motivation, and deeper linguistic awareness. Excessive dependence on AI-generated output can also lead to a superficial understanding of target-language structures and weakened translation competence. Thus, effective pedagogical integration must involve deliberate balance—combining AI-assisted tasks with hands-on translation practice that cultivates human judgment, problem-solving, and creative adaptation.

Post-editing activities offer a promising model for achieving this balance. By asking students to evaluate, edit, and reflect on LLM-produced translations, instructors can encourage critical engagement and linguistic analysis rather than passive acceptance of machine output. However, teachers must recognize that because LLMs produce fluent and natural-sounding texts, errors are often subtle and difficult to detect. Identifying these errors now requires deeper semantic and pragmatic understanding rather than surface-level correction. Training students to navigate these challenges will therefore be essential for developing the next generation of translators capable of working effectively alongside AI technologies.

Viewed through the lens of Vygotsky's sociocultural theory, LLMs can function as scaffolding tools in translation education by operating within learners' Zone of Proximal Development, where a more capable model provides support just beyond their current competence (Vygotsky, 1978). Because LLMs can approximate human translation quality and offer broadly acceptable feedback on general texts, they can serve as such a "more capable other," enabling students to imitate, compare, and refine their own output through guided interaction rather than passive dependence. At the same time, LLMs can foster learner autonomy in line with Self-Regulated Learning (SRL), which emphasizes learners' ability to plan, monitor, and evaluate their performance (Zimmerman, 2002). When used in low-stakes or formative tasks, they allow students to test hypotheses, revise drafts, and track improvement in real time, thereby strengthening metacognitive awareness and self-assessment. Instructors, however, should design in-class activities where students first translate manually to protect core linguistic and problem-solving skills. They can then use tools such as ChatGPT in take-home or lab tasks to generate or refine translations, write brief reflections on where the model helped or failed, and engage in collective post-editing of LLM-produced texts. Asking students to compare their own translations with LLM output, identify discrepancies, and justify preferred solutions can help them develop critical awareness of AI tools while consolidating the linguistic and strategic competences required for independent, informed translation performance.

FUNDING

This work was funded and supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [KFU254141].

REFERENCES

[1] 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. https://doi.org/10.1080/23311983.2025.2508031

[2] Aghai, M. (2024). ChatGPT vs. Google Translate: Comparative Analysis of Translation Quality. *Iranian Journal of Translation Studies*, *22*(85). Retrieved November 2, 2025, from https://dorl.net/dor/20.1001.1.17350212.1403.22.1.9.2

[3] AlAfnan, M. A. (2024). Large Language Models as Computational Linguistics Tools: A Comparative Analysis of ChatGPT and Google Machine Translations. *Journal of Artificial Intelligence and Technology*, *5*, 20–32. https://doi.org/10.37965/jait.2024.0549
[4] Al-khresheh, M. H., & Almaaytah, S. A. (2018). English Proverbs into Arabic through Machine Translation. *International Journal of Applied Linguistics and English Literature*, *7*(5), 158–166. https://doi.org/10.7575/aiac.ijalel.v.7n.5p.158
[5] Al Rousan, R., Jaradat, R., & Malkawi, M. (2025). ChatGPT translation vs. human translation: An examination of a literary text. *Cogent Social Sciences*, *11*(1). https://doi.org/10.1080/23311886.2025.2472916
[6] Ataman, D., Birch, A., Habash, N., Federico, M., Koehn, P., & Cho, K. (2025). Machine translation in the era of large language models: A survey of historical and emerging problems. *Information*, *16*(9). Retrieved November 2, 2025, from https://www.mdpi.com/2078-2489/16/9/723
[7] Alghamdi, F. A., & Alotaibi, H. (2025). Using AI in Translation Quality Assessment: A Case Study of ChatGPT and Legal Translation Texts. *Electronics*, *14*(19). https://doi.org/10.3390/electronics14193893
[8] Almaaytah, S. A., & Alzobidy, S. A. (2023). Challenges in rendering Arabic text to English using machine translation: A systematic literature review. *IEEE Access*, *11*, 94772–94779. Retrieved November 2, 2025, from https://ieeexplore.ieee.org/abstract/document/10233872/
[9] Alkhofi, A. (2024). The Use of Google Translate in the Arabic-English Classroom. *Theory and Practice in Language Studies*, *14*(12), 3861-3870.
[10] Alkhofi, A. (2025). Can ESL instructors spot machine translation? Evidence from the Arabic-English classroom. In *Forum for Linguistic Studies* (Vol. 7, pp. 340-350).
[11] Alzain, E., Nagi, K. A., & Algobaei, F. (2024). The Quality of Google Translate and ChatGPT English to Arabic Translation: The Case of Scientific Text Translation. *Forum for Linguistic Studies*, *6*(4), 837–849. Retrieved November 2, 2025, from https://www.academia.edu/download/121385964/The_Quality_of_Google_Translate_and_ChatGPT_English_to_Arabic_Translation_The_Case_of_Scientific_Text_Translation.pdf
[12] Bahdanau D, Cho K, & Bengio Y. (2014). Neural machine translation by jointly learning to align and translate. In: *Proceedings of the 3rd International Conference on Learning Representations*.
[13] Bowker, L. (2019). Machine translation literacy as a social responsibility. *Proceedings of the Language Technologies for All (LT4All)*, 104–107. Retrieved November 2, 2025, from https://lt4all.elra.info/media/papers/O7/145.pdf
[14] Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J., ... & Roossin, P. S. (1990). A statistical approach to machine translation. *Computational linguistics*, *16*(2), 79-85.
[15] Cai, L. (2024). *How does ChatGPT Compare with Conventional Neural Machine Translation: Ingenta Connect*. Retrieved November 2, 2025, from https://www.ingentaconnect.com/content/plg/jts/2024/00000004/00000001/art00003
[16] Calvo-Ferrer, J. R. (2024). Can you tell the difference? A study of human vs machine-translated subtitles. *Perspectives*, *32*(6), 1115–1132. https://doi.org/10.1080/0907676X.2023.2268149
[17] Cao, S., & Zhou, T. (2025). Exploring the Efficacy of ChatGPT-Based Feedback Compared With Teacher Feedback and Self-Feedback: Evidence From Chinese-English Translation. *Sage Open*, *15*(3). https://doi.org/10.1177/21582440251369204
[18] Chan, V., & Tang, W. K.-W. (2024). GPT and Translation: A Systematic Review. *2024 International Symposium on Educational Technology (ISET)*, 59–63. https://doi.org/10.1109/ISET61814.2024.00021
[19] Corpas Pastor, G., & Noriega-Santiáñez, L. (2024). Human versus Neural Machine Translation Creativity: A Study on Manipulated MWEs in Literature. *Information (Switzerland)*, *15*(9). https://doi.org/10.3390/info15090530
[20] Constantine, P. (2020). Literary Translation Pedagogy in the United States: New Trends. *Translation Review*, *106*(1), 10–14. https://doi.org/10.1080/07374836.2019.1625833
[21] Dong, D., Wu, H., He, W., Yu, D., & Wang, H. (2015). Multi-task learning for multiple language translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers, pp. 1723–1732). Retrieved November 2, 2025, from https://aclanthology.org/P15-1166.pdf
[22] 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. https://doi.org/10.34190/ejel.23.2.4006
[23] Duan, H., Gao, X., & Zhang, Y. (2025). The Application of AI Translation Tools in Improving Students' Translation Fidelity and Accuracy. *Arab World English Journal*, *16*, 290–306. https://doi.org/10.24093/awej/AI.16
[24] Ed-Dali, R. (2025). Assessing DeepSeek R1 and ChatGPT 4.5 in Arabic-English literary translation: Performance, challenges, and implications. *Cogent Arts & Humanities*, *12*(1). https://doi.org/10.1080/23311983.2025.2531183
[25] El-Saadany, M. R. (2024). A Comparative Study between Chat GPT and Human Translation in Translating English Proverbs into Arabic. *مجلة البحث العلمي في الآداب*, *25*(5), 24–54. https://doi.org/10.21608/jssa.2024.257874.1592
[26] Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: Human translation vs. AI large language models (Gemini and ChatGPT). *Cogent Social Sciences*, *10*(1), 2410998. https://doi.org/10.1080/23311886.2024.2410998
[27] Gao, Y., Wang, R., & Hou, F. (2024). How to Design Translation Prompts for ChatGPT: An Empirical Study. *Proceedings of the 6th ACM International Conference on Multimedia in Asia Workshops*, 1–7. https://doi.org/10.1145/3700410.3702123
[28] Gjorevski, A., Li, M., & Cox, T. L. (2025). Exploring the Potential of ChatGPT for Evaluating English Essays in a Criterion-Based Assessment. *TESOL Quarterly*, *59*(S1), S251-S279. https://doi.org/10.1002/tesq.70011
[29] Habib, R., Alkhawaja, L., Khoury, O., & Al-Sayyed, S. (2025). Six NMT Systems, One Language Pair: Which Best Translates Arabic-English? *World Journal of English Language*, *16*(1). https://doi.org/10.5430/wjel.v16n1p1
[30] Haider, A. S., & Alkhatib, R. (2024). Subtitling English Legal Acronyms into Arabic: Human vs Machine. *Kutafin Law Review*, *11*(4). Retrieved November 2, 2025, from https://kulawr.msal.ru/jour/article/view/424
[31] He, Y. (2021). Challenges and Countermeasures of Translation Teaching in the Era of Artificial Intelligence. *Journal of Physics: Conference Series*, *1881*(2). https://doi.org/10.1088/1742-6596/1881/2/022086

[32] Huang, Y., Li, D., & Cheung, A. K. F. (2025). Evaluating the linguistic complexity of machine translation and LLMs for EFL/ESL applications: An entropy weight method. *Research Methods in Applied Linguistics*, *4*(3). https://doi.org/10.1016/j.rmal.2025.100229
[33] Jiang, Z. (2025). Does LLM translation align with translation universals? A cross-genre simplification study on English-Chinese translation based on dependency grammar. *PLOS ONE*, *20*(6). https://doi.org/10.1371/journal.pone.0324830
[34] Jiang, Z., & Zhang, Z. (2024). *Can ChatGPT Rival Neural Machine Translation? A Comparative Study*. CoRR. Retrieved November 2, 2025, from https://openreview.net/forum?id=pxNZz3EfES
[35] 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*, 1–21. https://doi.org/10.1080/1750399X.2025.2533074
[36] Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator? Yes with GPT-4 as the engine.* arXiv Preprint arXiv:2301.08745. Retrieved November 2, 2025, from https://arxiv.org/abs/2301.08745
[37] Jia, Y., Carl, M., & Wang, X. (2019). Post-editing neural machine translation versus phrase-based machine translation for English–Chinese. *Machine Translation*, *33*(1), 9–29.
[38] Johnston, H., Wells, R. F., Shanks, E. M., Boey, T., & Parsons, B. N. (2024). Student perspectives on the use of generative artificial intelligence technologies in higher education. *International Journal for Educational Integrity*, *20*(1). https://doi.org/10.1007/s40979-024-00149-4
[39] Koponen, M. (2016). Is machine translation post-editing worth the effort? A survey of research into post-editing and effort. *The Journal of Specialised Translation*, *25*(2), 131–148.
[40] Krüger, R. (2023). Some reflections on the interface between professional machine translation literacy and data literacy. *Journal of Data Mining & Digital Humanities* (*IV. Challenges for professional translation)*. Retrieved November 2, 2025, from https://jdmdh.episciences.org/9728
[41] 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
[42] Lee, S.-M. (2023). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning*, *36*(1–2), 103–125. https://doi.org/10.1080/09588221.2021.1901745
[43] Liu, F. (2018). Ways to improve effect of college English translation teaching. *International Conference on Education, Psychology, and Management Science*, 979–983. Retrieved November 2, 2025, from https://webofproceedings.org/proceedings_series/ESSP/ICEPMS%202018/ICEPMS208.pdf
[44] Lu, S.-C., Xu, C., Kaur, M., Edelen, M. O., Pusic, A., & Gibbons, C. (2025). Can machine translation match human expertise? Quantifying the performance of large language models in the translation of patient-reported outcome measures (PROMs). *Journal of Patient-Reported Outcomes*, *9*(1), 94. https://doi.org/10.1186/s41687-025-00926-w
[45] Manapbayeva, Z., Zaurbekova, G., Ayazbekova, K., Kazezova, A., & Pirmanova, K. (2024). AI in Literary Translation: ChatGPT-4 vs. Professional Human Translation of Abai's Poem 'Spring.' *Procedia Computer Science*, *251*, 526–531. https://doi.org/10.1016/j.procs.2024.11.143
[46] Mohsan, M., & Nayab, D. e. (2024). Estimating and Comparing Translation Skills: A Comparative Study of ChatGPT and Human Translation. *Journal of Development and Social Sciences*, *5*(3), 75–86. https://doi.org/10.47205/jdss.2024(5-III)08
[47] Moneus, A. M., & Sahari, Y. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, *10*(6). https://doi.org/10.1016/j.heliyon.2024.e28106
[48] Mizumoto, A., & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics*, *2*(2), 100050. Retrieved November 2, 2025, from https://www.sciencedirect.com/science/article/pii/S2772766123000101
[49] Ng, S.-L., & Ho, C.-C. (2025). Generative AI in Education: Mapping the Research Landscape Through Bibliometric Analysis. *Information*, *16*(8), 657. https://doi.org/10.3390/info16080657
[50] OpenAI. (2022, November 30). *Introducing ChatGPT*. OpenAI https://openai.com/index/chatgpt/
[51] Özmat, D., & Akkoyunlu, B. (2024). Artificial Intelligence-Assisted Translation in Education: Academic Perspectives and Student Approaches. *Participatory Educational Research*, *11*(H. Ferhan Odabaşı Gift Issue), 151–167. https://doi.org/10.17275/per.24.99.11.6
[52] Povilaitienė, M., & Kasperė, R. (2022). Machine Translation for Post-Editing Practices. *Scientific Journal of Mykhailo Dragomanov State University of Ukraine. Series 9. Current Trends in Language Development*, *24*, 47–62. https://doi.org/10.31392/NPU-nc.series9.2022.24.04
[53] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., & Brennan, S. E. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, *372*. Retrieved November 2, 2025, from https://www.bmj.com/content/372/bmj.n71.short
[54] Rao, P., McGee, L. M., & Seideman, C. A. (2024). A Comparative assessment of ChatGPT vs. Google Translate for the translation of patient instructions. *Journal of Medical Artificial Intelligence*, *7*. https://doi.org/10.21037/jmai-24-24
[55] Rico, C., & Gonzalez Pastor, D. (2022). The role of machine translation in translation education: A thematic analysis of translator educators' beliefs. *Translation & Interpreting-the International Journal of Translation and Interpreting*, *14*(1), 177–197. https://doi.org/10.12807/ti.114201.2022.a010
[56] Rico Pérez, C. (2024). *Re-thinking machine translation post-editing guidelines*. Retrieved November 2, 2025, from https://docta.ucm.es/entities/publication/36ab888d-39bb-400d-bbf1-e379ed394296
[57] Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: A systematic review. *Language Resources and Evaluation*, 56(2), 593–619. https://doi.org/10.1007/s10579-021-09537-5
[58] Sadiq, S. (2025). Evaluating English-Arabic translation: Human translators vs. Google Translate and ChatGPT. *Journal of Languages and Translation*. https://doi.org/10.21608/jltmin.2025.423147
[59] Sanz-Valdivieso, L., & López-Arroyo, B. (2023). Google Translate vs. ChatGPT: Can non-language professionals trust them for specialized translation. In *International Conference Human-Informed Translation and Interpreting Technology (HiT-IT 2023)* (pp. 97–107).

[60] Shormani, M. Q., & Alfahad, A. (2025). Artificial Intelligence or Human: The Use of ChatGPT in the Academic Translation for Religious Texts. *SAGE Open*, *15*(3). https://doi.org/10.1177/21582440251343954
[61] Stapleton, P., & Kin, B. L. K. (2019). Assessing the accuracy and teachers' impressions of Google Translate: A study of primary L2 writers in Hong Kong. *English for Specific Purposes*, *56*, 18–34. Retrieved November 2, 2025, from https://www.sciencedirect.com/science/article/pii/S0889490619300158
[62] Su, Y., Xu, S., & Liu, K. (2025). Adapt or adopt? Examining the efficacy of ChatGPT in providing translation feedback. *The Interpreter and Translator Trainer*, 1–21. https://doi.org/10.1080/1750399X.2025.2541486
[63] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems* (Vol. 2, pp. 3104-3112). MIT Press
[64] Tafa, T. O., Hashim, S. Z. M., Othman, M. S., Alhussian, H., Nasser, M., Abdulkadir, S. J., Huspi, S. H., Adeyemo, S. O., & Bena, Y. A. (2025). *Machine Translation Performance for LowResource Languages: A Systematic Literature Review*. IEEE Access. Retrieved November 2, 2025, from https://ieeexplore.ieee.org/abstract/document/10972018/
[65] Tavares, C., Oliveira, L., Duarte, P., & Da Silva, M. M. (2023). Artificial intelligence: A blessing or a threat for language service providers in Portugal. *Informatics*, *10*(4), 81. Retrieved November 2, 2025, from https://www.mdpi.com/2227-9709/10/4/81
[66] Toledo-Báez, C. (2024). Post-editing and human-machine parity in neural machine translation: An empirical study from professional translation. *Lebende Sprachen*, *69*(2), 434–463. https://doi.org/10.1515/les-2024-0003
[67] Vieira, L. N. (2019). Post-editing of machine translation. In *The Routledge handbook of translation and technology* (pp. 319–336). Routledge. Retrieved November 2, 2025, from https://www.taylorfrancis.com/chapters/edit/10.4324/9781315311258-22/post-editing-machine-translation-lucas-nunes-vieira
[68] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (Vol. 86). Harvard University Press.
[69] Wang, H., Wu, H., He, Z., Huang, L., & Church, K. W. (2022). Progress in Machine Translation. *Engineering*, *18*, 143–153. https://doi.org/10.1016/j.eng.2021.03.023
[70] Wang, Y. (2023). Artificial Intelligence Technologies in College English Translation Teaching. *Journal of Psycholinguistic Research*, *52*(5), 1525–1544. https://doi.org/10.1007/s10936-023-09960-5
[71] Weaver, W. (1952). Translation. In *Proceedings of the Conference on Mechanical Translation*. Retrieved November 2, 2025, from https://aclanthology.org/1952.earlymt-1.1.pdf
[72] Wu, J. (2023). A comparative analysis of Chinese-English translation quality based on ChatGPT: A case study of Chinese characteristic words. *Journal of Social Science Humanities and Literature*, *6*(5), 53–58. https://www.adwenpub.com/index.php/jsshl/article/view/71
[73] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., … & Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*. arXiv. https://doi.org/10.48550/arXiv.1609.08144
[74] Woodrum, C. (2024). ChatGPT and Language Translation A Small Case Study Evaluating English—Mandarin Translation. In H. Degen & S. Ntoa (Eds.), *Artificial Intelligence in Hci, Pt Iii, Ai-Hci 2024* (Vol. 14736, pp. 147–157). Springer International Publishing Ag. https://doi.org/10.1007/978-3-031-60615-1_10
[75] Xu, S., Su, Y., & Liu, K. (2025). Investigating student engagement with AI-driven feedback in translation revision: A mixed-methods study. *Education and Information Technologies*, *30*(12), 16969–16995. https://doi.org/10.1007/s10639-025-13457-0
[76] Yamada, M. (2019). The impact of Google neural machine translation on post-editing by student translators. *The Journal of Specialised Translation*, *31*(1), 87–106. Retrieved November 2, 2025, from https://www.researchgate.net/profile/Masaru-Yamada/publication/364994185_art_yamada_newpdf/data/63620fc037878b3e87755cf7/art-yamada-new.pdf
[77] Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice*, *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2
[78] Yang, Y., Liu, R., Qian, X., & Ni, J. (2023). Performance and perception: Machine translation post-editing in Chinese-English news translation by novice translators. *Humanities and Social Sciences Communications*, *10*(1). https://doi.org/10.1057/s41599-023-02285-7
[79] 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. https://doi.org/10.1080/1750399X.2025.2535239
[80] Zhang, H. (2022). Comparison between Human Translation and Machine Translation in Translating the Publicity Text of Haihunhou Museum. In *2022 8th Annual International Conference on Network and Information Systems for Computers* (ICNISC) (pp. 177–180). https://doi.org/10.1109/ICNISC57059.2022.00045
[81] Zhang, W., Li, A. W., & Wu, C. (2025). University students' perceptions of using generative AI in translation practices. *Instructional Science*, *53*(4), 633–655. https://doi.org/10.1007/s11251-025-09705-y
[82] Zhang, Z., Abdullah, S. N. S., Abdullah, M. A. R., & Zhou, L. (2025). Google Translate or ChatGPT-4? A Multi-Metric Evaluation of Chinese-to-English Technical Translation. *Forum for Linguistic Studies*, *7*(9), 770–788. https://doi.org/10.30564/fls.v7i9.11014

Anas M. Alkhofi was born in Saudi Arabia on January 1988. He received a BA in English Language from King Faisal University, KSA, in 2009, an MA in TESOL from the University of Central Florida, USA, in 2015, and a PhD in Educational Linguistics from the University of New Mexico, USA, in 2020. He began his career as an English teacher in 2010 and then served as an Assistant Professor in the English Department at Imam Muhammad University from 2010 to 2022. He was the Head of the English Department from 2021 to 2023. Currently, he is an Assistant Professor in the English Department at King Faisal University in Hufuf, KSA. Previous publications include Alkhofi, (2025) Man vs. Machine: Can AI Outperform Student Translations? and Alkhofi, (2024) The Use of Google Translate in the Arabic-English Classroom. He is involved in research concerning vocabulary acquisition, lexical access, and the integration of genAI in English teaching and learning.
https://orcid.org/0009-0000-5097-0126