Translation and Interpreting Research
Vol. 2, No. 5, March 2025, 45-58
tir.atu.ac.ir
DOI: 10.22054/tir.2025.86625.1044

A Comparison of AI-Assisted, AI-Revised and Human-Scaffolded Translations in ESP Classes

Mahboobeh Khosravani — English Language Department, University of Birjand, Birjand, Iran
Sajjad Farokhipour* — Department of Language and Literature, Shahid Mahallati University of Islamic Sciences, Qom, Iran
Davoud Padiz — Department of English Language and Literature, Yasuj University, Yasuj, Iran
Bamshad Hekmatshoartabari — Department of English Language and Literature, Yasuj University, Yasuj, Iran

Abstract

AI-assisted translation has gained increasing attention in recent years, yet its effectiveness remains underexplored. The present study sought to shed light on the role of AI (ChatGPT) in mediating translation. To this end, 46 postgraduate ESP students majoring in three sub-disciplines of politics (across three classes) were selected through convenience sampling. No outliers were identified in these classes, and each was assigned to one experimental group (AI-assisted group, N = 16; AI–Human Revised group, N = 16; Human-only Scaffolded group, N = 14). A posttest-only control group design was adopted, and each group was mediated according to its respective intervention protocol. AI was instructed to follow a graduated mediation protocol developed for the purposes of this study. The final translations were evaluated both qualitatively and quantitatively. Findings revealed that the end product of the AI-assisted group, compared with the human-involved groups, exhibited major translation deficiencies ranging from the lexico-semantic level to syntax, the syntax–semantic interface, and rhetorical patterns. Additional procedural deficiencies were also observed and reported. Furthermore, participants' translations were assessed using a rubric, and quantitative analysis showed that both human-involved groups significantly outperformed the AI-only group.

Keywords: AI, ESP, feedback, GPT, mediation, prompt, translation

Original Article
Accepted: 17.03.2025
Received: 30.02.2025

*Corresponding author: eltcmu@gmail.com

Cite this article: Khosravani, M., Farokhipour, S., Padiz, D., & Hekmatshoartabari, B. (2025). A comparison of AI-assisted, AI-revised and human-scaffolded translations in ESP classes. *Translation and Interpreting Research*, *2*(5), 45-58. DOI: 10.22054/tir.2025.86625.1044

Publisher: ATU Press

Translation and Interpreting Research is the journal of Research Institute for Translation Studies (RITS), affiliated with Allameh Tabataba'i University, Tehran, Iran.

1. Introduction

A surge in use of English as a lingua franca on one hand and its users' demand for a greater latitude in employing tools on the other, has opened favorable opportunity for AI-assisted language education around the world (Maurice Gayed et al., 2025). Besides, Hwang et al. (2020) maintain that AI-based language processing has achieved an unparalleled position in applied linguistics research. Similarly, Farokhipour et al. (2025) have underscored that AI-based models of analyzing linguistic data have attracted researchers due to the theoretical and practical spread of these models in recent years. Machine translation and AI-based translation have also been a focal point in this development. This development has been so critically fast that Zhang and Zhang (2019) expressly hold that machine transition would fundamentally shift the role of human translators and degrade its position to an only post-edition agents at most. Consistent with these transitions of theories and transformations of roles, a huge bulk of research is carried out on different dimensions of this phenomena such as comparing machine and human translation (e.g. Lu et al., 2023; Abdelhalim et al., 2025), machine translation methods (e.g. Jiao et al., 2024; Karmaker & Feng, 2023) and more recently neural machine translation (e.g. Lee, 2023) and prompt engineering of translation machines or neural networks (e.g. Peng et al., 2023).

One of the most applied AI-based translation machines that is largely discussed in education discipline is chatGPT which due to its capability in emulating human cognitive capability (Godwin-Jones, 2022) and its enormous capacity in text processing and text production (Siu, 2023) is largely employed in education research (Su et al., 2023; Jacob et al., 2024; Werdiningsih et al., 2024). This tool has been potent and cogent enough to be used widely for language education in particular (Al-Rousan et al., 2025) due to its peculiar capacity in processing human interaction and appropriate feedback provision (Surameery & Shakor, 2023), its rootedness in large human linguistic data and models (Deng & Lin, 2023) and its immense competence in text paraphrasing and interpretation (Kalla & Smith, 2023). For the same reasons, it is also widely used in translation research and practice as a machine mediator by giving prompts for text interpretation (Gu, 2023), translation quality assessment tool (Kocmi & Federmann, 2023) and inspiring human translators (Jiao et al., 2023). Though there exist considerable conflicting views regarding accuracy of chatGPT translation (e.g. Hendy et al., 2023; Farokhipour et al., 2025) the studies submitting proof on the reliability and worth of this version of AI translation win through the literature. Despite this promising capacity, the use of chatGPT in language teaching is an unexplored area especially in low AI-resourced languages such as Persian. An embryonic area of language research which can find benefit in chatGPT is ESP texts' interpretation and translation as both a text-decoding tool and a feedback provider. Thus the current research is an innovative attempt to explore the effect of AI-assisted and AI-revised language interpretation compared with human assisted language interpretation on final Persian translated text among ESP post-graduate students of politics enrolled in an ESP course.

2. Literature Review

The way literature on artificial intelligence is related to the current research can be discussed from three different backgrounds; the studies inspecting the role of AI in education and language learning in particular, the researches delving into the role of AI in translation and last but not least, the studies recollecting the insights and experiences associated with AI and feedback provision in language class. These dimensions are looked back on briefly below.

AI and Language Education

Deeply rooted in large language models and previously established human patterns and frameworks, generative artificial intelligence tools are equipped with enough resources to promote language learning in various ways including enriching and refining teaching and learning experiences, provision of instant individualized prompts and feedbacks, facilitation of comprehension and text decoding and interpretation (Evmenova et al., 2024). The same findings were already arrived at by Borup (2023). Guan et al. (2025) also replicated these findings and concluded that AI's capability and facility in promoting language education has brought it to the focal point of research in the discipline mostly due to creation of dynamic and interactive learning environment. Being fine-tuned to learning conditions and individual language learners' needs is one more advantage of AI arrived at by Fathi et al (2024). Besides, from a psychological perspective, review of literature indicated that inclusion of AI technology in language class result in higher language learning and speaking willingness, lower levels of anxiety and higher levels of engagement and enjoyment (Chen, 2024; Guan et al., 2024). Doshi and Hauser (2023), too, proved the effect of AI in promoting language learners' creativity. Furthermore, a number of current studies have arrived at promising findings for inclusion of AI in curriculum development and syllabus design for personalized content and material which in turn foster learning and achievement (e.g. Lozano & Fontao, 2023). And last but not least, a scant review of literature reveals the propitious outlooks for inclusion of AI in special language education (Marino et al., 2023; Koraishi, 2023). Despite these favorable findings, however, some infelicitous findings are also associated with using AI in education in general and language education in particular including the increased possibility of cheating (Trust et al., 2023), inclusion of errors in provided feedbacks and responses to language learners (Borji, 2023), reduced thinking, especially critical thinking habits (Rezende Junior & López-Simó, 2024) and teachers' negative attitude toward AI in elimination of human from language learning (Rahimi et al., 2025) as well as seize of job opportunities (Hopcan et al., 2024).

AI and Translation/Interpretation

A scant review of literature in the domain of AI and text interpretation and translation reveals that AI-based machine translation has resulted in a new discipline in translation theory and practice which is at odds with traditional human-translation from a variety of perspectives (Yang, 2022; Moneus & Sahari, 2024). Despite that, a significant number of studies hold a contrary view and assert that machine translation, compared with human translation, is half-grown and largely immature (e.g. Li et al., 2020; Farokhipour et al., 2025). Despite these conflicting views, recently published studies have submitted positive and confirming evidences on efficacy of AI in teaching and learning translation and text interpretation (Al-Rousan et al., 2025). While Larroyed (2023) and Chow et al., (2024) stand firm behind the reliability, precision and value of AI-based machine translation as comparable with human translation, Siu (2023) sheds light on the faithful transmission of intended meaning from one language to another by AI-based translation. Consistently, Peng et al. (2023) concluded that, even in the case of largely distant source-target languages, AI is proved as a proficient translator. Even, Wang et al. (2024) arrived at a surprising end result that AI translation outperforms human translation in the sub-discipline of commercial discourse translation. Similarly, Karabayeva and Kalizhanova (2024) found that AI translators can convey literary stylistic devices such as metaphors, similes, imagery and tone from one language to another accurately and proficiently. Khoshafah (2023), too, submitted further evidence on the capacity of AI for accurate transfer of genre from one language to another. However, one more strand of studies on AI translation implicitly evinces at least partial inadequacy of AI translation and suggests some modifications for its improvement in areas such as prompts, resources, patterns, procedures, and decoding methods. (e.g. Gu, 2023). Yet, a group of more critical studies have raised concern regarding the accuracy of AI translation, at least in certain types of texts (e.g., Hendy et al., 2023; Deng & Lin, 2023).

AI and Feedback Provision

AI-generated feedback is reported facilitative for language learning since it amends mistakes based on given prompts (Rezai et al. 2024), amending essays and promoting discourse elements (Tate et al., 2023) and reducing cognitive and emotional loads of language learners (Zhou et al., 2023). Yet, Xu et al. (2025) hold that, despite the experimentally and theoretically established weight of feedback in language education, this role in AI-based translation education is an unexplored area. Concerning this subject, Su et al. (2023) and Banihashem et al. (2024) have postulated that AI-based technology is potent and cogent enough to provide tuned, detailed and effective feedback on language learner's performance. Though almost all this studies have explored effect of feedback on learning in areas other than translation, Ekin (2023) posits that AI-based feedback can also be employed in translation education too. Besides, Herbold et al. (2023) had already proved that AI can be used in language translation education successfully. Notwithstanding, exploring the effect of AI-generated feedback on translation is still in its infancy.

Therefore, the current research is an attempt to blend insights from the above-explored areas of research and adopt a proper research design to investigate the effect of AI-generated feedback and AI-generated/human-revised feedback on translation quality of ESP learners compared with human-independent dynamic feedback in Iranian context.

3. Methodology

Context and Design

Literature review showed that only a scant number of studies have investigated the effect of AI on translation in real educational setting. Therefore, the present research was carried out to highlight a new dimension of this contribution in an unexplored area of ESP texts interpretation and translation in Iranian context. Considering Shan (2022) which underscores the dialectical pluralism and richness of results in mixed-methods research, mixed-methods design was selected aimed at comparing AI-assisted, AI-revised and human-scaffolded translations on translation quality of ESP learners. First, a qualitative comparison of the translations carried out with the selected approaches was conducted. Then, adopting a quantitative quasi-experimental approach and using a quantitative estimation rubric, the effect of using the selected approaches on translation quality was estimated.

Sampling

To fulfill the goals of the research, two different sources of data were targeted. The primary source of data that used in early phase of research for qualitative comparison was the final Persian translated scripts of ESP language learners. However, for the quasi-experimental phase of the study 46 postgraduate students in three different disciplines of political sciences were selected through convenience sampling. These students were enrolled in an ESP course in which a similar content (English for students of politics I & II by Hormoz Davarpanah) was taught. These learners were then randomly assigned to AI-assisted group (N=16) receiving only instant AI feedback, AI-Human Revised group (N=16), receiving AI-modified feedback and human-only scaffolded group (N=14) receiving normative dynamic intervention using a tuned mediation inventory. The homogeneity of the participants was already reckoned. No significant outlier was identified.

Instruments

Consistent with the objectives of the research, three major instruments were employed in this study that are described below:

A. AI-Guided Mediation Protocol

Though AI automatically provides graduated instant prompts for language learners and translators, to avoid direct translation and keeping adherence to the pedagogical worth and values of dynamic intervention, a mediation protocol was devised by the researcher and given to ChatGPT4 to be used as mediatory protocol. This protocol is depicted in Table 1.

Table 1. AI-Guided Mediation Protocol Moves

Error Category	Mediation Move
Ambiguity	Clarification request
Tone mismatch	Reformulation prompt
Meaning failure	Semantic/morphologic awareness
Syntactic inconsistency	Grammatical reflection
Structural inadequacy	Contrastive analysis
Repeating error	Noticing
Unprecise translation	Precision recast
Functional mismatch	Contextual clues
Rhetorical patterns	Genre awareness

B. Glocalized Mediation Inventory

AI-Mediated groups received AI feedbacks on their performance, the human-mediated group, however, was scaffolded through a glocalized mediation inventory adopted from Farokhipour (2019) that is proved more tuned with Iranian language learning context, compared with standardized mediation inventories which entailed 12 moves arranged from implicit to explicated as depicted in Table 2.

Table 2. Glocalized mediation inventory used for scaffolding translation

Mediation
Ask for modification
Refusing answer
Recalling instruction
Pinpointing erroneous segment
Narrow down error site
Pinpointing the type of error
Providing metalinguistic clues
Providing similar examples
Providing binary options
Providing correct form
Teaching the rule
Reviewing learning

C. Translation Quality Estimation Rubric

In order to quantify the performance of the participants on translation task, and building on translation quality assessment models, a weighted rubric was developed that entailed seven criteria in which the total score is calculated on a 20-point scale. This rubric is shown in Table 3.

Table 3. Translation Quality Assessment Rubric

Criterion	Weight	Score (1–4)	Subtotal
Accuracy of Meaning	25%		
Terminology Use	20%		
Grammar & Syntax	15%		
Style & Tone	10%		
Cohesion & Coherence	10%		
Rhetorical Patterns	10%		
Discourse Metafictions	10%		

Final Score Calculation: (Sum of Weighted Scores) × 7 = Score out of 20

Procedure

In order to compare the effects of different mediation types on the translation quality of ESP learners, three groups were created and a posttest-only control group design was used, in which the three groups of students were given their appropriate feedback. A similar translation task containing 317 words was selected from the aforesaid coursebook, and the participants in each group were asked to translate the text from English into Persian. The first group only received AI mediation, while the second group received AI mediation scaffolded by the human agent. However, the third group received a normative graduated mediation by a human agent only. To make ChatGPT follow our developed instruction (AI-Guided Mediation Protocol Moves) and avoid direct correct translations, a custom GPT template was used and uploaded wherein AI persistently followed our mediation protocol. To avoid full-text translation by AI, participants were given the text sentence by sentence rather than the whole text. Each group's performance was evaluated independently and rated both qualitatively and quantitatively. The results of the qualitative juxtaposition of participants' translations were presented. Furthermore, in order to weigh the differences between the performances of the groups from a quantitative viewpoint, the translations were measured against a rubric and their mean scores were compared by one-way analysis of variance in SPSS (Version 23).

4. Results

Qualitative Findings

As it was designated above, the primary goal of the current study was qualitative juxtaposition of translation quality in the final product of AI-assisted, AI-revised and human-scaffolded translation groups. Elaborating on AI-only feedback deficiencies, the results of this phase of the research are put adjacent in Table 4 with the following examples.

Profound Semantic Understanding: The word "Code" in most translations of the AI-only assisted group was translated as "دستورالعمل" or "رمز", while its accurate translation is "مجموعه قوانین".

Global Consistency: The term "state" in most translations of AI-only assisted group was translated differently across the text as "حکومت", "دولت", or "نظام", while its overall meaning across the whole text was "دولت".

Syntactic Ambiguity: "WH" question makers which were used as connectors were translated as interrogative questions which resulted in Persian syntactic ambiguity (for instance, "what" was wrongly translated as "چه چیزی" instead of "چیزی که").

Syntactic shift resulting in ambiguity. For instance in the sentence "law passed by the legislators representing the will of the electorate" is ambiguously translated as " قانونی توسط قانونگذاران تصویب شد تا نشانگر اراده حوزه انتخابیه باشد".

Other ambiguous sentences such as misinterpreting physical setting with political context, failure in determining agent, and over-expansion of structure leading to over-translation were observed that resulted in syntactic ambiguity.

Syntax-Semantic Interface: Misuse of a verb type realizing a specific meta-function residing in syntax-semantic interface. For instance, "He was realized as an ethical thinker" is translated as " او به عنوان یک متفکر اخلاقی فهم شده بود" while the verb "realize" here is not classified as verb groups entailing the meaning of "understanding" to fulfill a mental process. Rather, "realize" is classified in a verb group denoting the meaning of "being" and fulfills a relational process that in Persian equals "به شمار آمدن".

Observing Tone: The verb 'dismissed' in "He dismissed all these propositions" is translated as either "رد کرد" or "نادیده گرفت" while the more enhanced tone is "فاقد اعتبار خواند".

Register Match: Sometimes AI suggestions were too simplistic or too journalistic and distant from political science register. For instance the phrase "built on legitimate political authority" is translated as "قدرت سیاسی مشروع" while more academic register is "مشروعیت اقتدار سیاسی". Another example is in the phrase "Euphrates civilization is both site and agent of civil regulation", where the word "site" is translated as "مکان" while the more accurate equivalent is "بستر".

Aligned Word Knowledge: Some technical words are sometimes mistaken by AI. For instance, the word "utility" which in political sciences denotes "فایده گرایی" is wrongly interpreted as "کارایی".

Accountable Translation: Use of arbitrary translations for one fixed term was evident without any specific rationale. For instance the word "governance" was arbitrarily translated as "حکومت" and "حکومت داری" while haphazardly capturing the more accurate equivalent, i.e. "حکمرانی" as well.

Table 4. Results of Qualitative Analysis of Translations

Error Type	AI-Assisted	AI-Revised	Human-Scaffolded
Profound semantic understanding	✗	✓	✓
Global consistency	✗	✓	✓
Syntactic ambiguity	✗	✓	✓
Syntax-semantic interface	✗	✓	✓
Observing tone	✗	✓	✓
Register match	✗	✓	✓
Aligned word knowledge	✗	✓	✓
Accountable translation	✗	✓	✓

In addition to the differences revealed above, it was also observed that AI-only translation feedback varies from AI-revised and human-only feedback in a number of dimensions including personalization and individual differences (while AI only stimulates monotonously for all leaners, human mediated feedbacks are more personalized and tuned to learners needs), emotional sensitivity (while AI is only responsive to concrete aspects of mediation, human mediation shows due consideration for human emotions through tuning tone and demonstrating empathy), error/mistake diagnosis (while AI feedback is systematic, human feedback is consistent with responsiveness moves of individuals and

shows considerations to ZPD of language learners and therefore is able to diagnose not only the nuances but also tell apart mistake from error and provide more consistent mediation), and more engagement with human mediator.

Quantitative Findings

In addition to the differences revealed above, it was also observed that AI-only translation feedback varies from AI-revised and human-only feedback in a number of dimensions, including personalization and individual differences (while AI only stimulates monotonously for all learners, human-mediated feedback is more personalized and tuned to learners' needs), emotional sensitivity (while AI is only responsive to concrete aspects of mediation, human mediation shows due consideration for human emotions through tuning tone and demonstrating empathy), error/mistake diagnosis (while AI feedback is systematic, human feedback is consistent with responsiveness moves of individuals and shows consideration of the ZPD of language learners and therefore is able to diagnose not only the nuances but also tell apart mistakes from errors and provide more consistent mediation), and greater engagement with the human mediator.

Table 5. ANOVA Results for Estimation of Differences in Translation Performance

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	295.932	2	147.966	64.635	.000
Within Groups	98.438	43	2.289		
Total	394.370	45			

Table 5 presents the results of the analysis of variance (ANOVA) conducted to examine differences in the mean scores of the three groups. The analysis revealed a significance value of 0.000 (i.e., $p < 0.001$), which is below the 0.05 threshold, indicating a statistically significant difference in mean scores among the groups participating in the translation task. To identify the specific group differences, a post hoc multiple comparison test was subsequently performed (Table 6).

Table 6. Tukey-HSD Post-Hoc Test for Multiple Comparison

(I) Groups	(J) Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
AI-Only	AI-Human	-4.937*	.535	.000	-6.24	-3.64
	Human-Only	-5.687*	.554	.000	-7.03	-4.34
AI-Human	AI-Only	4.938*	.535	.000	3.64	6.24
	Human-Only	-.750	.554	.374	-2.09	.59
Human-Only	AI-Only	5.688*	.554	.000	4.34	7.03
	AI-Human	.750	.554	.374	-.59	2.09

As it is indicated by table 6, there is a statistically significant difference in scores between the AI-only group and AI-Human group ($p = 0.000$), as well as between AI-only group and Human-only group ($p = 0.00$). However it was revealed that the difference between AI-Human group and human-only group is not significant ($p = 0.374$). In sum, there was a statistically significant difference between groups as determined by one-way ANOVA ($F(2, 43) = 64.635$, $p = 0.000$). A Tukey post-hoc test revealed that wherever human feedback is involved, the performance of the participants in translation task is promoted significantly compared with AI-only group.

5. Discussion

Kohnke et al. (2023) argue that large training datasets, the capacity to respond immediately to users' prompts in a conversational manner, and the ability to fine-tune feedback to learners' needs have given AI a distinct role in assisting language learning. Furthermore, many studies on the adequacy of advanced models of machine translation have shown that translations based on large language models such as ChatGPT perform well in rendering both the meaning and form of different languages (see Calvo-Ferrer, 2023; Jiao et al., 2023; and Skobo & Petricevic, 2023, among others). In addition, a substantial body of research (e.g., Yang et al., 2023; Kasneci et al., 2023; and Devlin et al., 2018) strongly supports the role of large language models such as ChatGPT in enhancing productive language skills such as speaking and writing. However, the findings of the current research revealed that this effect is difficult to generalize to translation. These results are consistent with Li (2024), who raised serious concerns regarding the ethical and technical issues of machine translation, such as the potential for cheating and the challenges of prompt engineering. The current findings also support Liu et al. (2024), who highlighted AI's failure to capture the nuances of language learners' performance.

Moreover, compared with other global languages, Persian is considered a low-resource language, and given the fact that AI translation models are highly dependent on such resources, AI-based English-to-Persian translation is vulnerable to a number of quality deficiencies, such as syntactic and semantic asymmetry. The findings in this respect are in line with Siu (2023), who concluded that AI translations into low-resource languages result in severe syntactic asymmetry. Inconsistent and non-accountable translations were predominantly observed in the performance of the AI-only supported group. This outcome can possibly be associated with the type of feedback provided by AI. Su et al. (2023) observed that vague AI feedback directly leads to inconsistent cognitive performance among AI users, which in turn results in non-accountable translations that undermine the global coherence of the target text.

The significant outperformance of human-involved groups over the AI-only group is also consistent with the systematic review conducted by Ali Mohsen et al. (2023), which revealed that without in-depth human revision, AI translations suffer from a variety of technical problems. With respect to rhetorical patterns, the findings of the current study also support Farokhipour et al. (2025), who concluded that machine translation, particularly ChatGPT, fails to convey many discourse-level elements of language, including ideational metafunctions, at least in certain texts.

In sum, the findings of the current research revealed that AI-only feedback is not conducive to highly adequate translation and faces numerous difficulties at all levels of language, from lexico-semantic elements to syntax, the syntax-semantic interface, and discourse-level rhetorical patterns. Although some studies have shown that incorporating language-specific add-on lexicons can mitigate many of these deficiencies (e.g., Liu & Zhu, 2023), it appears that final human revision of the translation remains the most reliable option.

6. Conclusion

This study pursued two primary objectives. First, it aimed to identify the differences between AI–Human modes of feedback on translation quality from a qualitative perspective. To this end, the final products of the translators were thoroughly evaluated, and points of difference were derived at various levels of language, from the word level to the discourse level. It was revealed that both groups receiving human-mediated translations outperformed the AI-only feedback group due to more profound semantic understanding, global consistency, syntactic clarity, the ability to convey

functions at the syntax–semantic interface, observance of tone, register matching, aligned word knowledge, and accountable translation.

The second purpose of the research was to investigate the differences between AI–Human modes of feedback on translation quality from a quantitative perspective. To this end, a posttest control group design was adopted, and a rating rubric was utilized to assess students' translations on a 20-point scale. A one-way analysis of variance was carried out to compare the means. A statistically significant difference was reported in the mean scores between the AI-only group and the AI–Human group ($p = 0.000$), as well as between the AI-only group and the Human-only group ($p = 0.000$). However, it was revealed that the difference between the AI–Human group and the Human-only group was not significant ($p = 0.374$). In sum, there was a statistically significant difference between groups as determined by one-way ANOVA ($F(2, 43) = 64.635$, $p = 0.000$). Furthermore, a Tukey post hoc test revealed that whenever human feedback was involved, the performance of participants in the translation task improved significantly compared with the AI-only group.

These findings have considerable pedagogical implications for EFL practitioners and stakeholders. The primary implication is for language teachers, who could consider integrating AI into the interpretation and translation of ESP texts. In addition, AI proved flexible enough to provide advanced prompts to support ESP learners, which in turn sheds more light on the viability of AI in dynamic assessment, mediating cognition, and fostering the internalization of learning. It was also shown that AI assistance in translation can reduce mediation time, as prompts are given instantly, which implies that AI mediation can inform alternative language instruction and assessment methods. Additionally, these findings revealed that AI cannot serve as an adequate mediator for translation on its own. The fact that translation accuracy still depends heavily on human cognition, emotion, and behavior provides a basis for modifying AI translation systems by designing specific prompt-tags to address these gaps.

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 and Humanities, 12*(1). https://doi.org/10.1080/23311983.2025.2508031

Ali Mohsen, M., Althebi, S., & Albahooth, M. (2023). A scientometric study of three decades of machine translation research: Trending issues, hotspot research, and co-citation analysis. *Cogent Arts & Humanities, 10*(1). https://doi.org/10.1080/23311983.2023.2242620

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

Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education, 21*(1), 23. https://doi.org/10.1186/s41239-024-00455-4

Borji, A. (2023). A categorical archive of ChatGPT failures. *arXiv*. *2302.03494.* https://doi.org/10.48550/arXiv.2302.03494

Borup, J. (2023). This was written by a human: A real educator's thoughts on teaching in the age of ChatGPT. *Educause Review*.

Calvo-Ferrer, J. R. (2023). 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

Chen, Y. C. (2024). Effects of technology-enhanced language learning on reducing EFL learners' public speaking anxiety. *Computer Assisted Language Learning, 37*(4), 789–813.

Chow, J. C. L., Wong, V., & Li, K. (2024). Generative pre-trained transformer-empowered healthcare conversations: Current trends, challenges, and future directions in large language model-enabled medical chatbots. *BioMedInformatics, 4*(1), 837–852.

Deng, J., & Lin, Y. (2023). The benefits and challenges of ChatGPT: An overview. *Frontiers in Computing and Intelligent Systems, 2*(2), 81–83. https://doi.org/10.54097/fcis.v2i2.4465

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint. 1810.04805v2*. https://doi.org/10.48550/arXiv.1810.04805

Doshi, A. R., & Hauser, O. (2023). Generative artificial intelligence enhances creativity. *SS RN*. https://doi.org/10.2139/ssrn.4535536

Ekin, S. (2023). Prompt engineering for ChatGPT: A quick guide to techniques, tips, and best practices. *TechRxiv preprint*. https://doi.org/10.36227/techrxiv.22683919.v2

Evmenova, A. S., Borup, J., & Joan Kang Shin. (2024). Harnessing the power of generative AI to support all learners. *TechTrends*. https://doi.org/10.1007/s11528-024-00966-x

Farokhipour, S., Padiz, D., & Hekmatshoartabari, B. (2025). Investigating the accuracy of GPT's translation in transferring syntax-semantic interface: A case study of ideational meta-functions in English translation of Quran. *Language Research*, *15*(3), 1-25. http://doi:10.22059/jolr.2025.391037.666912

Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System, 121*, 103254. https://doi.org/10.1016/j.system.2024.103254

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers and Education: Artificial Intelligence, 3*, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. *Language Learning Technology, 26*(2), 5–24. https://hdl.handle.net/10125/73474

Guan, L., Zhang, Y., & Gu, M. M. (2024). Examining generative AI mediated informal digital learning of English practices with social cognitive theory: A mixed method study. *ReCALL*. https://10.1017/S0958344024000259

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*, 29, 8435-8463. https://doi.org/10.1007/s10639-023-12146-0

Han, C., & Lu, X. (2023). Can automated machine translation evaluation metrics be used to assess students' interpretation in the language learning classroom? *Computer Assisted Language Learning, 36*(5–6), 1064–1087. https://doi.org/10.1080/09588221.2021.1968915

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., wadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *arXiv preprint*, 2302.09210 https://doi.org/10.48550/arXiv.2302.09210

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports, 13*(1), 18617. https://doi.org/10.1038/s41598-023-45644-9

Hopcan, S., Türkmen, G., & Polat, E. (2024). Exploring the artificial intelligence anxiety and machine learning attitudes of teacher candidates. *Education and Information Technologies, 29*(6), 7281–7301. https://doi.org/10.1007/s10639-023-12086-9

Hwang, G.-J., Xie, H., Wah, B., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence, 1*(1), 100001. https://doi.org/10.1016/j.caeai.2020.10000

Jacob, S. R., Tate, T., & Warschauer, M. (2024). Emergent AI-assisted discourse: A case study of a second language writer authoring with ChatGPT. *Journal of China Computer-Assisted Language Learning, 4,* 1-22. https://doi.org/10.1515/jccall-2024-0011

Jiao, H., Peng, B., Zong, L., Zhang, X., & Li, X. (2024). Gradable chat GPT translation evaluation. *ArXiv.org*. https://arxiv.org/abs/2401.09984

Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? A preliminary study. *arXiv preprint*, 2301-08745.

Kalla, D., & Smith, N. (2023). Study and analysis of ChatGPT and its impact on different fields of study. *International Journal of Innovative Science and Research Technology, 8*, 827–833. https://doi.org/10.5281/zenodo.7767675

Karabayeva, I., & Kalizhanova, A. (2024). Evaluating machine translation of literature through rhetorical analysis. *Journal of Translation and Language Studies, 5*(1), 1–9. https://doi.org/10.48185/jtls.v5i1.962

Karmaker, S, & Feng, D. (2023). TELeR: A general taxonomy of LLM prompts for benchmarking complex tasks. *ArXiv (Cornell University)*. https://doi.org/10.18653/v1/2023.findings-emnlp.946

Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 1–9. https://doi.org/10.1016/j.lindif.2023.102274

Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy. *Research Square Online First*, 1–20. https://doi.org/10.21203/rs.3.rs-2814154/v1

Kocmi, T., & Federmann, C. (2023). Large language models are state-of-the-art evaluators of translation quality. https://doi.org/10.48550/arXiv.2302.14520

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Chat GPT for language teaching and learning. *RELC Journal, 54*, 537–550. https://doi.org/10.177/00336882231162868

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology, 3*(1), 55–72. https://bit.ly/43en7el

Larroyed, A. (2023). Redefining patent translation: The influence of ChatGPT and the urgency to align patent language regimes in Europe with progress in translation technology. *GRUR International, 72*(11), 1009–1017. https://doi.org/10.1093/grurint/ikad099

Li, B., Liu, H., Wang, Z., Jiang, Y., Xiao, T., Zhu, J & Li, C. (2020). Does multi-encoder Help? A Case study on context-aware neural machine translation, *arXiv preprint*. ArXiv: 2005-03393.

Liu, S., & Zhu, W. (2023). An analysis of the evaluation of the translation quality of neural machine translation application systems. *Applied Artificial Intelligence, 37*(1). https://doi.org/10.1080/08839514.2023.2214460

Liu, G. L., Darvin, R., & Ma, C. (2024). Exploring AI-mediated informal digital learning of English (AI-IDLE): A mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer Assisted Language Learning*, 1–29. https://doi.org/10.1080/09588221.2024.2310288

Lozano, A., & Fontao, C. B. (2023). Is the education system prepared for the irruption of artificial intelligence? A study of the perceptions of students of primary education degree from a dual perspective: Current pupils and future teachers. *Education Sciences, 13*(7), 733. https://doi.org/10.3390/educsci13070733

Marino, M. T., Vasquez, E., Dieker, L., Basham, J., & Blackorby, J. (2023). The future of artificial intelligence in special education technology. Journal of Special Education Technology, 38(3), 404–416. https://doi.org/10.1177/01626434231165977

Moneus, A. M., & Sahari, Y. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon, 10*(6), e28106. https://doi.org/10.1016/j.heliyon.2024.e28106

Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Yuaxin, O., & Tao, D. (2023). Towards making the most of ChatGPT for machine translation. https://doi.org/10.2139/ssrn.4390455

Rahimi, A. R., Sheyhkholeslami, M., & Pour, A. M. (2025). Uncovering personalized L2 motivation and self-regulation in ChatGPT-assisted language learning: A hybrid PLSSEM- ANN approach. *Computers in Human Behavior Reports, 17*, Article 100539. https://doi.org/10.1016/j.chbr.2024.100539

Rezai, A., Namaziandost, E., & Hwang, G. J. (2024). How can ChatGPT open promising avenues for L2 development? A phenomenological study involving EFL university students in Iran. *Computers in Human Behavior Reports, 16*, 100510. https://doi.org/10.1016/j.chbr.2024.100510

Rezende Junior, M. F., & López-Simó, V. (2024). What are the perceptions of physics teachers in Brazil about ChatGPT in school activities? *Journal of Physics: Conference Series, 2693*(1), 12011. https://doi.org/10.1088/1742-6596/2693/1/01201.1

Shan, Y. (2022). Philosophical foundations of mixed methods research. *Philosophy Compass, 17*(1). https://doi.org/10.1111/phc3.12804

Siu, S. C. (2023). ChatGPT and GPT-4 for professional translators: Exploring the potential of large language models in translation. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.4448091

Škobo, M., & Petričević, V. D. (2023). Navigating the challenges and opportunities of literary translation in the age of AI: Striking a balance between human expertise and machine power. Društvene i humanističke Studije, 8(2), 317–336. https://doi.org/10.51558/2490-3647.2023.8.2.317

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing, 57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Surameery, N. M. S., & Shakor, M. Y. (2023). Use ChatGPT to solve programming bugs. International *Journal of Information Technology and Computer Engineering, 3*(31), 17–22.

Tate, T. P., Doroudi, S., Ritchie, D., Xu, Y., & Uci, M. W. (2023). Educational research and AI-generated writing: Confronting the coming Tsunami. Preprint in EdArXiv. https://doi.org/10.35542/osf.io/4mec3

Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education, 23*(1), 1–23.

Wang, Y., and A. J. Daghigh. (2024). Effect of text type on translation effort in human translation and neural machine translation post-editing processes: Evidence from eye-tracking and keyboard-logging. *Perspectives 32* (5): 961–976.

Werdiningsih, I., Marzuki, R. D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities, 11*(1), 2392388.

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*. https://doi.org/10.1007/s10639-025-13457-0

Yang, C. (2022). The application of Artificial Intelligence in translation teaching. *Proceedings of the 4th International Conference on Intelligent Science and Technology*, 56-60.

Zhang, W., & Zhang, E. (2019). RETRACTED: Application of neural network machine translation in college translation teaching. *International Journal of Emerging Technologies in Learning (IJET), 14*(19), 16. https://doi.org/10.3991/ijet.v14i19.10690

Zhou, L., Dewaele, J. M., Lochtman, K., & Xi, Y. (2023). Foreign language peace of mind: A positive emotion drawn from the Chinese EFL learning context. *Applied Linguistics Review, 14*(5), 1385–1410. https://doi.org/10.1515/applirev-2021-0080