KSU-JRLT

Journal of Research in Language & Translation
Volume 6, Special Issue (December 2026)

جامعة الملك سعود
King Saud University

إعادة النظر في مفهوم كفاية المترجم في عصر الذكاء الاصطناعي التوليدي: نحو إطار للترجمة التشاركية بين الإنسان والذكاء الاصطناعي

إبراهيم الأسمري
قسم الترجمة ، جامعة الملك خالد
أبها، المملكة العربية السعودية
ORCID: https://orcid.org/0000-0003-2323-3176
ifasmary@kku.edu.sa

Received: 25 October 2026; Revised: 15 November 2026; Accepted: 25 November 2026

DOI: 10.33948/JRLT-KSU-6-S-5

الملخص

تهدف هذه الدراسة إلى إعادة النظر في مفهوم "كفاية المترجم" في ضوء التحولات التي أحدثها توظيف الذكاء الاصطناعي التوليدي في ممارسة الترجمة، وذلك باستخدام تصميم بحثي تقاربي يجمع بين الأساليب الكمية والنوعية، ويتألف من ثلاثة مسارات بحثية: أولها مراجعة نقدية لإطارين مرجعيين من أطر كفايات المترجمين وهما: إطار "باكتي" (PACTE Group, 2003, 2020) وإطار "إ إم تي" (EMT Expert Group, 2009, 2022)، وتهدف هذه المراجعة النقدية إلى بحث الجوانب التي تستلزم مزيدًا من التطوير في هذين الإطارين حول الكفايات التي تتطلبها مهنة الترجمة عند توظيف أنظمة الذكاء الاصطناعي التوليدي ونماذجه الحديثة. وثاني تلك المسارات دراسة مسحية شملت ثمانية وتسعين متر جماً من الممارسين والأكاديميين وطلبة الدارسات العليا في الترجمة، وذلك لاستقصاء تصوراتهم بشأن أهمية تلك الكفايات المرتبطة بالذكاء الاصطناعي ومدى استيعابها في البرامج التدريبية والأكاديمية. أما المسار الثالث فقائم على مدونة لغوية متوازية تضم خمسة وأربعين نصاً باللغة الإنجليزية، يقابل كلاً منها ترجمتان إلى العربية، إحداهما ترجمة بشرية منشورة أنجزها مترجم محترف، والأخرى أنتجها نموذج الذكاء الاصطناعي التوليدي GPT-4o. وتشتمل المدونة على 572,652 كلمة موزعة على أربع سجلات لغوية (قانونية، وعلمية، وأدبية، وصحفية)، جرى تحليلها باستخدام مقياسَي BLEU وchrF، ومقياسَي TTR وMATTR، وتحليل الكلمات المفتاحية في اتجاهَي المقارنة. وتشير نتائج الدراسة إلى تزايد أهمية المهام التقييمية والإشرافية للمترجم إلى جانب مهامه التقليدية في إنتاج النصوص. وقد صنّف المشاركون كفاية التقييم النقدي بوصفها الأكثر أهمية، وأشارت استجاباتهم إلى وجود فجوات بين الأهمية المتصورة للكفايات ومدى توفر التدريب عليها. كما كشف تحليل المدونة عن فروق بين الترجمات البشرية والترجمات المولَّدة بالذكاء الاصطناعي في التنوع المعجمي، والتعبيرات المرتبطة بالسجل اللغوي، والتعامل مع أسماء الأعلام والاختصارات، في حين أظهرت المقاييس الآلية تفاوتًا بين السجلات اللغوية في درجات التشابه مع الترجمات البشرية المرجعية. واستنادًا إلى هذه النتائج والمراجعة النظرية، تقترح الدراسة إطار كفايات المترجم في بيئة العمل التشاركي بين الإنسان والذكاء الاصطناعي، والموسوم بإطار "هاكتك" (HACTC)، والذي يضم ست كفايات موزعة على مجموعتين وظيفيتين، هما الكفايات التشغيلية المرتبطة بالذكاء الاصطناعي والكفايات النقدية التقييمية، إضافة إلى كفاية أخلاقية سابعة تتداخل مع المجموعتين. ويهدف الإطار المقترح إلى تطوير الأسس التي يقوم عليها إطارا PACTE وEMT، من خلال تقديم تصور أكثر تحديدًا للمعارف والمهارات التي تتطلبها الترجمة القائمة على التعاون بين الإنسان والذكاء الاصطناعي.

الكلمات المفتاحية: دراسات الترجمة القائمة على المدونات اللغوية؛ الترجمة من الإنجليزية إلى العربية؛ الذكاء الاصطناعي التوليدي؛ إطار HACTC؛ الترجمة التشاركية بين المترجم والذكاء الاصطناعي؛ كفاية المترجم؛ إعداد المترجمين.

KSU-JRLT

Journal of Research in Language & Translation
Volume 6, Special Issue (December 2026)

جامعة الملك سعود
King Saud University

Reconceptualising Translation Competence in the Age of Generative AI: Towards a Human–AI Collaborative Framework

Ibrahim Alasmri

Translation Department, King Khalid University
Abha, Saudi Arabia

ORCID: https://orcid.org/0000-0003-2323-3176
ifasmary@kku.edu.sa

Received: 25 October 2026; Revised: 15 November 2026; Accepted: 22 November 2026

DOI: 10.33948/JRLT-KSU-6-S-5

Abstract

This study investigates the reconfiguration of translator competence in response to the increasing integration of generative artificial intelligence (GenAI) into translation practice. It employs a convergent mixed-methods design comprising three complementary components. The first critically examines the PACTE Group (2003, 2020) and EMT Expert Group (2009, 2022) frameworks to identify areas requiring further development in relation to GenAI-mediated translation. The second uses a survey of 98 professional translators, translation academics, and postgraduate translation students to examine the perceived importance of AI-related competencies and the availability of relevant training. The third employs a parallel corpus of 45 English source texts, each paired with a professionally published human Arabic translation and a GPT-4o-generated translation. Comprising 572,652 words across legal, scientific, literary, and journalistic registers, the corpus is analysed using BLEU, chrF, TTR, MATTR, and bidirectional keyword analysis. The findings suggest an increasing emphasis on the translator's evaluative and supervisory responsibilities alongside established text-production skills. Survey respondents rated Critical Evaluation as the most important competency area and reported gaps between competency importance and training availability. The corpus analysis revealed differences in lexical diversity, register-sensitive expressions, and the treatment of proper nouns and acronyms, while automated evaluation showed variation in similarity to human reference translations across registers. Drawing on these findings and the theoretical review, the study proposes the Human–AI Collaborative Translation Competence (HACTC) Framework. The framework comprises six competencies organised into two functional groups, AI Operational and Critical-Evaluative, with Ethical Competence as a seventh, cross-cutting component. It extends the foundations of PACTE and EMT by providing a more explicit account of the knowledge and skills required for human–AI collaborative translation.

Keywords: corpus-based translation studies; English–Arabic translation; generative artificial intelligence; HACTC framework; human–AI collaborative translation; translator competence; translator training

Introduction

Translation competence has been the subject of considerable theoretical attention within translation studies over the past three decades, resulting in the development of several well-established theoretical and empirical frameworks (Prieto Ramos, 2024). Models such as PACTE Group (2003, 2020) and the European Master's in Translation framework (EMT Expert Group, 2009, 2022) have made important contributions to understanding the different components of translator competence, distinguishing professional competence from bilingual proficiency, and providing a basis for translator training and professional standards. However, these frameworks do not fully account for the emerging forms of human–AI collaboration associated with the increasing integration of generative artificial intelligence (GenAI) into professional translation practice. This development therefore raises questions about the extent to which existing models can adequately account for the changing role of translators and the competencies required in AI-mediated translation environments.

Earlier translation technologies, including computer-assisted translation (CAT) tools and neural machine translation systems, have already contributed to significant changes in translation practice, particularly in relation to translator productivity, workflow, and post-editing (Koponen, 2016; O'Brien, 2012). GenAI systems, however, are increasingly capable of performing complex linguistic and communicative tasks traditionally undertaken by human translators. These capabilities vary considerably, and expert human translation continues to demonstrate advantages in quality-sensitive, domain/register-specific, and culturally complex tasks, particularly in terms of document-level coherence, cultural mediation, and stylistic judgement (Yan et al., 2024). More importantly, the increasing use of GenAI is changing the role of translators from primary producers of translated texts towards evaluators, post-editors, and supervisors of AI-generated translations. This shift has important implications for how translator competence is conceptualised, operationalised, and taught (Prieto Ramos, 2024).

Existing research on AI-assisted translation has not yet fully addressed these developments at the theoretical level. Research within computational linguistics and natural language processing has largely focused on comparing the quality of GenAI-generated translations with conventional machine translation systems, using automatic evaluation metrics alongside human evaluation and error analysis (Hendy et al., 2023; Jiao et al., 2023; Vilar et al., 2023). Other studies have investigated specific aspects of GenAI-assisted post-editing, including its effects on translation quality and the cognitive and technical effort involved (Kwok et al., 2025; Yao et al., 2025). Although recent research has begun to reconsider translator competence in AI-mediated environments (Prieto Ramos, 2024), a comprehensive theoretical framework that accounts for human–AI collaboration as an integral component of professional translation competence remains insufficiently developed.

Against this background, the current study aims to address this theoretical gap by investigating how the integration of GenAI into translation practice is changing the competencies required of professional translators, and by proposing a revised framework of translator competence that accounts for these developments. The study employs a convergent mixed-methods design consisting of three components: a critical review of existing translation competence frameworks; a survey of professional translators, academics, and advanced translation students (N = 98); and a corpus-based comparison of GenAI-generated translations and professionally published human translations across four registers.

The remainder of the paper is organised as follows. The Literature Review examines existing translation competence models and relevant research on GenAI-mediated translation. The Methodology section presents the research questions and outlines the three components of the study. The Results section presents the survey and corpus findings, followed by the proposed Human–AI Collaborative Translation Competence (HACTC) Framework. The

Discussion section examines these findings in relation to the research questions and previous literature. Finally, the Conclusion and Recommendations section considers the implications for translator training and professional practice and identifies directions for future research.

Literature Review and Theoretical Background

Translation Competence: Definitions and Models

The PACTE model (PACTE Group, 2003, 2020) and the European Master's in Translation framework (EMT Expert Group, 2009, 2022) have made important theoretical, empirical, and institutional contributions to understanding translation competence. However, their capacity to account for the competencies required in GenAI-mediated translation practice warrants further examination.

The PACTE model, developed by the PACTE Group at the Autonomous University of Barcelona, identifies five interconnected sub-competencies: bilingual, extra-linguistic, knowledge about translation, instrumental, and strategic, alongside psycho-physiological components (PACTE Group, 2003). The strategic sub-competency is assigned a central coordinating function, responsible for planning, monitoring, and evaluating the translation process and activating the other sub-competencies as required. The PACTE Group's subsequent empirical research, consolidated in its 2020 publication, provided further support for the model and the importance of strategic competence in distinguishing translation expertise from bilingual proficiency. Earlier empirical research by the PACTE Group (2005) examined translation decision-making and the use of internal and external resources, identifying differences between professional translators and other language professionals. More recently, Hazaea and Qassem (2024) drew on PACTE's instrumental sub-competence to investigate its representation in translator training programmes at Saudi universities.

The EMT framework, developed by the European Commission's Directorate-General for Translation, provides a professional reference for translator training programmes (EMT Expert Group, 2009, 2022). Unlike PACTE, which primarily addresses the cognitive and empirical dimensions of translation competence, EMT focuses on the knowledge and skills required in professional practice. Its successive revisions have increasingly recognised the importance of translation technologies, including machine translation and post-editing. The 2022 framework identifies five competence areas and incorporates technological knowledge and skills relevant to contemporary translation practice. Previous research has examined the EMT framework from both empirical and educational perspectives. Chodkiewicz (2012) investigated how professional translators and translation students perceived the relevance of its competence areas, while Krause (2024) examined the applicability of the 2022 framework in relation to evolving professional requirements. However, although these developments acknowledge the changing technological environment, the framework does not systematically address how collaboration with GenAI may affect the relationships between the different components of translator competence. This limitation is particularly relevant as translators increasingly assume responsibilities for evaluating and revising AI-generated translations.

Generative AI in Translation

The increasing integration of GenAI into translation practice has attracted considerable research attention, particularly regarding its effects on translation quality and professional workflows. A growing body of computational research has compared the performance of large language models with conventional neural and commercial machine translation systems, using automatic evaluation metrics alongside human evaluation and error analysis (Hendy et al., 2023; Jiao et al., 2023; Vilar et al., 2023). Other studies have extended these comparisons to

professional human translation across different languages, domains, and levels of translator expertise (Yan et al., 2024).

The findings of these studies vary considerably. Yan et al. (2024) found that GPT-4 produced error rates comparable to those of junior translators, while more experienced professional translators generally demonstrated better performance. Earlier research on neural machine translation also showed that differences between human and machine translation become more apparent when evaluation considers document-level context rather than isolated sentences (Läubli et al., 2018). Similarly, Karpinska and Iyyer (2023) demonstrated that providing document-level context can improve LLM-based literary translation, although limitations concerning critical errors and stylistic consistency remain. These findings suggest that linguistic fluency alone is insufficient to establish translation quality, particularly where contextual interpretation, cultural knowledge, and stylistic judgement are required.

Within the Arabic translation context, several studies have investigated the quality of GenAI-generated translations in comparison with human translation and conventional MT systems. Khoshafah (2023) found that ChatGPT could produce usable translations of relatively straightforward texts, but encountered difficulties in preserving contextual, historical, and cultural nuances. Alzain et al. (2024) compared Google Translate and ChatGPT in English-to-Arabic scientific translation, with professional evaluation and error analysis favouring Google Translate. Similarly, Sahari et al. (2024) reported that human translators were more successful than GPT-3 in conveying figurative and idiomatic meanings. Taken together, these findings provide further evidence of the limitations of GenAI-generated translations in contexts requiring cultural and contextual interpretation.

Although this body of research contributes to understanding the capabilities and limitations of GenAI in translation, much of it has focused on evaluating translation output rather than investigating the competencies required of translators working with these

technologies. Consequently, the implications of GenAI for professional translation competence remain less systematically investigated.

Human–AI Collaboration as an Emerging Professional Model

The increasing use of GenAI in translation practice has contributed to changes in the role of professional translators, particularly in workflows where AI systems produce initial translations that are subsequently evaluated and revised by human translators. This development may be viewed as an emerging form of human–AI collaborative translation, in which professional expertise is increasingly exercised through evaluating, revising, and validating AI-generated output.

This perspective builds on earlier research addressing collaboration and technology in translation practice. Jiménez-Crespo (2017) examined the changing organisation of translation through online collaborative practices, while O'Brien (2012) considered translation as a form of human-computer interaction. Kiraly's (2000) social constructivist approach to translator education also emphasises the development of professional competence through collaboration and engagement in authentic translation practices. Although these approaches were developed in different technological contexts, they provide a theoretical basis for examining collaboration as an important dimension of translator competence.

In GenAI-mediated translation environments, the translator's responsibilities may extend beyond producing translated texts to evaluating linguistic accuracy, cultural appropriateness, stylistic consistency, and the reliability of AI-generated output. This requires the ability to identify translation problems, assess the limitations of AI systems, formulate effective prompts, and make informed decisions about accepting or revising generated translations. These requirements overlap with established translation competencies, particularly strategic, instrumental, and evaluative competence, but also introduce new

demands associated with the use of generative technologies. Consequently, the integration of GenAI raises questions about the relative importance of existing competencies and the need to identify additional competencies that support effective human–AI collaboration.

Research Gap

The literature reviewed above indicates that research on translation competence and GenAI-assisted translation has developed from different theoretical and methodological perspectives. Established competence models provide a basis for understanding the knowledge and skills required of professional translators, while recent studies of GenAI have largely focused on translation quality, system performance, and specific aspects of AI-assisted translation practice. Although existing models recognise the importance of technological and strategic competence, the implications of human–AI collaboration for the structure and relative importance of different translation competencies remain insufficiently investigated.

Furthermore, studies evaluating GenAI translation performance have identified linguistic, contextual, and cultural limitations that require human evaluation and intervention (Khoshafah, 2023; Sahari et al., 2024; Yan et al., 2024). However, although previous research has examined GenAI translation performance across languages and domains (Yan et al., 2024) and reconsidered translator competence in response to technological developments (Prieto Ramos, 2024), the relationship between the textual characteristics of GenAI-generated translations and the competencies required for effective human–AI collaboration remains insufficiently investigated. This therefore remains an area where further theoretical and empirical investigation is required, particularly in identifying how existing competencies may need to be extended or reorganised to account for human–AI collaborative translation.

Methodology

This study employs a convergent mixed-methods design consisting of three complementary components: a critical review of existing translation competence models, a mixed-methods survey, and a corpus-based comparison of human and GenAI-generated translations. Together, these components provide a theoretical and empirical basis for examining the changing requirements of translator competence and developing the proposed Human–AI Collaborative Translation Competence (HACTC) Framework. The study addresses the following four research questions:

RQ1: How is the integration of generative AI reshaping the professional role and identity of translators in contemporary practice?

RQ2: What competencies are considered important for effective human–AI collaboration in professional translation, based on participants' perceptions, the theoretical review, and corpus-based findings?

RQ3: To what extent do existing translation competence frameworks, principally PACTE and EMT, adequately account for the demands of human–AI collaborative practice?

RQ4: What are the implications of the findings of the present study for translator training curricula and professional benchmarking?

Component One: Critical Review

The first component consists of a critical review of translation competence models, particularly PACTE Group (2003, 2020) and EMT Expert Group (2009, 2022), alongside recent research on AI-assisted translation and human–AI collaboration. The review aims to identify the extent to which existing frameworks account for the knowledge and skills required in GenAI-mediated translation practice. Relevant literature was identified through searches of Scopus, Google Scholar, and the ACL Anthology, using terms including translation

competence, generative AI, large language models, post-editing, human–AI collaboration, and machine translation evaluation. No date restriction was applied to translation competence literature to ensure the inclusion of foundational studies and the historical development of established competence models. In contrast, GenAI-related studies were limited to 2022–2025 to focus on recent developments following the emergence and widespread adoption of large language models in translation practice.

Component Two: Mixed-Methods Survey

The second component consists of a mixed-methods survey administered to 98 participants, including professional translators, translation academics, and advanced postgraduate translation students. Purposive sampling was used to include participants from different professional backgrounds, institutions, and translation specialisations. Participants were recruited through professional translation associations, university translation departments, and academic networks in Saudi Arabia and the broader Arab region.

The survey comprised three sections. The first collected information about participants' professional backgrounds, translation experience, specialisations, translation directions, and use of GenAI tools. The second used five-point Likert scales to assess the perceived importance and availability of training in six initially identified competency areas: AI Literacy, Prompt Engineering, Post-editing Competence, Critical Evaluation, Ethical Competence, and Cultural Awareness. These areas were identified before the survey and subsequently considered alongside the findings of the theoretical review and corpus analysis in developing the HACTC Framework. The third section included four open-ended questions addressing changes in professional roles, emerging competency requirements, translator training, and additional observations concerning GenAI.

Quantitative responses were analysed descriptively, with mean scores calculated for each competency in terms of perceived importance and training availability. Differences between these scores were used to identify potential gaps in current training provision.

Responses to the four open-ended questions were analysed thematically, drawing on Braun and Clarke's (2006) six-phase approach. This approach provides a systematic framework for identifying and interpreting recurring patterns of meaning across qualitative data. The analysis was primarily inductive, allowing themes to emerge from participants' responses rather than being determined exclusively by predefined competency categories. The resulting themes were subsequently examined in relation to the research questions and the competency areas identified in the theoretical review. This approach enabled the qualitative findings to complement the quantitative ratings by providing further insight into participants' perceptions of GenAI-related competencies and their implications for translator training.

Component Three: Parallel Corpus Analysis

The third component employs a parallel corpus consisting of 45 English source texts and their Arabic translations across four registers: literary, legal, scientific, and media. The English source texts and their professionally published Arabic translations were selected from the parallel component of a larger corpus previously used in Alasmri (2024). For the present study, 45 English source texts and their corresponding human Arabic translations were selected across four registers: literary, legal, scientific, and media. The texts were selected from published literary works, academic publications, news archives, and official legal documents, with the aim of maintaining comparability in text type and register. For the present study, each English source text was additionally translated using GPT-4o, resulting in a corpus of 135 texts comprising 572,652 words.

Table 1 presents the composition of the corpus. The Arabic translations contain fewer words than the English originals, a difference that may partly reflect morphological and orthographic characteristics of Arabic, particularly inflectional morphology and attached clitics (Holes, 2004; Habash et al., 2022). A comparable difference was reported by Al-Harthi et al. (2024), although such linguistic characteristics do not necessarily explain all differences in text length.

Table 1
Corpus Composition (Word Count Across Domains)

Corpus Component	Literary	Legal	Scientific	Media	Total
English originals	53,494	52,570	53,007	54,186	213,257
Human translations (HT)	44,092	47,100	46,386	50,265	187,843
AI translations (GPT-4o)	40,341	42,687	42,827	45,697	171,552
Register total	**137,927**	**142,357**	**142,220**	**150,148**	**572,652**

Note. Table 1 reports orthographic word counts. Token counts produced by corpus analysis software may differ because of tokenisation procedures, including the treatment of punctuation and other textual elements.

The AI translations were generated using GPT-4o (OpenAI, 2024) through the OpenAI API at a temperature setting of 0.3. A system prompt instructed the model to produce one Arabic sentence for each English source sentence to maintain sentence-level correspondence for subsequent comparison. This approach was intended to preserve the parallel structure required for automated translation evaluation and lexical analysis.

Corpus analysis comprised two complementary approaches. First, BLEU (Papineni et al., 2002) and chrF (Popović, 2015) were calculated using SacreBLEU (Post, 2018), with the professionally published human translations serving as the reference texts. BLEU was calculated using the standard 13a tokeniser, while chrF was calculated using the default SacreBLEU settings. Both metrics were applied separately to each of the 45 complete texts. Mean scores and standard deviations were then calculated for each register and for the complete set of texts. These document-level averages were used to examine differences in similarity to the human reference translations across registers.

Second, the human and AI translations were analysed using corpus-based lexical measures. Type-Token Ratio (TTR) was calculated to assess overall lexical diversity, while Moving-Average Type-Token Ratio (MATTR) was computed on raw words within individual files using a moving window of 100 words (Covington & McFall, 2010). The human and AI translation texts were also compiled into two separate corpora using Sketch Engine (Kilgarriff et al., 2014). Bidirectional Keyword Analysis was conducted, with each translation corpus serving as the reference corpus for the other. Keywords were ranked using the Simple Maths keyness score to identify lemmas occurring at higher relative frequencies in one corpus than in the other. The selected keywords were subsequently examined in context and categorised according to their linguistic functions, including discourse markers, modal expressions, stylistic and terminological choices, and proper-noun treatment.

The corpus findings were considered alongside the survey results and theoretical review to identify the competencies relevant to human–AI collaborative translation and to inform the development of the proposed HACTC Framework.

Ethical Considerations

Ethical approval was obtained from the Research Ethics Committee at King Khalid University (Approval No. ECM#2026-1903). Survey participation was voluntary, and respondents were informed of the study's purpose, the anonymity of their responses, and their right to withdraw before submission. No personally identifiable information was collected.

Results

This section presents the empirical findings of the study, drawing on the survey responses and comparative corpus analysis. The survey findings address participants' perceptions of GenAI-related competencies and training needs, while the corpus analysis examines differences between professionally published human translations and GPT-4o-

generated translations. These findings, considered alongside the critical review of existing competence frameworks, inform the development of the proposed Human–AI Collaborative Translation Competence (HACTC) Framework, which is also presented in this section.

Survey Findings

Participant Profile

The survey yielded 98 valid responses across four professional categories. Postgraduate students constituted the largest group (n = 35, 35.7%), followed by translation academics (n = 32, 32.7%), academics who also practised professional translation (n = 18, 18.4%), and professional translators (n = 13, 13.3%). Experience levels varied, with 33 participants reporting less than three years of experience (33.7%), 36 reporting 3–7 years (36.7%), 12 reporting 8–15 years (12.2%), and 17 reporting more than 15 years (17.3%). Thus, 29.6% had at least eight years of experience. Most participants reported translating in both Arabic–English directions (n = 77, 78.6%). The use of GenAI tools was also widespread, with 51.0% reporting daily use, 38.8% occasional use, and 10.2% having used these tools without regular adoption.

Competency Ratings: Importance and Training Availability

Participants rated six initially identified competency areas on two five-point scales measuring perceived importance and the availability of relevant training. Table 2 presents the mean scores, standard deviations, percentages of ratings of 4 or 5, and differences between the two dimensions.

Table 2

Importance and Training Availability Ratings for HACTC Competencies (N = 98)

Competency	Importance M (SD)	≥4 (%)	Training M (SD)	≥4 (%)	Gap
AI Literacy	3.77 (1.48)	62.2	3.05 (1.28)	36.7	0.71
Prompt Engineering	3.87 (1.40)	65.3	2.87 (1.27)	30.6	1.00

Competency	Importance M (SD)	≥4 (%)	Training M (SD)	≥4 (%)	Gap
Post-editing Competence	3.87 (1.59)	68.4	2.97 (1.29)	32.7	0.90
Critical Evaluation	3.91 (1.47)	70.4	2.86 (1.21)	27.6	1.05
Ethical Competence	3.77 (1.50)	64.3	2.99 (1.26)	34.7	0.78
Cultural Awareness	3.84 (1.56)	69.4	2.78 (1.29)	28.6	1.06

Note. Ratings were collected on a five-point Likert scale (1 = not at all; 5 = very much). ≥4 (%) represents the percentage of respondents assigning a rating of 4 or 5. Gap = mean importance rating minus mean training availability rating. Gaps were calculated from unrounded means.

Across all six competency areas, mean importance ratings exceeded mean training availability ratings, indicating a consistent perceived gap between professional requirements and available training. Critical Evaluation received the highest importance rating (M = 3.91), while Cultural Awareness recorded the largest training gap (1.06), followed closely by Critical Evaluation (1.05) and Prompt Engineering (1.00). These findings suggest that respondents considered these competencies particularly important while reporting comparatively limited opportunities for training in them.

Qualitative Thematic Analysis

Thematic analysis of the open-ended responses (Braun & Clarke, 2006) identified four main themes concerning the role of GenAI in translation, the continuing importance of human expertise, and the implications for translator training.

Theme 1: AI Capabilities and Limitations

Participants generally recognised the potential of GenAI to support translation practice while emphasising the need for human evaluation and supervision. The responses suggest that the usefulness of these technologies depends partly on translators' ability to identify their limitations and assess the appropriateness of their output.

Theme 2: Human Cultural and Pragmatic Competence

Cultural mediation and pragmatic interpretation were recurrent concerns in the responses. Participants referred to difficulties in handling culturally specific references, idiomatic expressions, implicit meanings, humour, and stylistic variation. One participant observed: “AI is still unable to grasp what lies behind the text, historical, religious, and local customs and traditions” (P62). Another participant highlighted the importance of contextual information in localisation: “Localization projects are collaborative, and translators work on small segments rather than the full text. When AI is applied to these segments, it may lack the broader cultural and functional context” (P32). These responses indicate that participants considered cultural and contextual interpretation important areas of human expertise in AI-mediated translation.

Theme 3: Implications for Translator Training

Participants identified several areas requiring greater attention in translator training, particularly prompt engineering, post-editing, critical evaluation, and the ethical use of GenAI. They also emphasised the importance of maintaining existing linguistic and translation competencies alongside the development of AI-related skills. One respondent stated: “It is essential to redesign many curricula to include new skills such as post-editing and critical evaluation, while continuing to build foundational translation skills, without which students cannot distinguish correct from incorrect” (P58).

Theme 4: Professional Responsibility and Human Oversight

A further theme concerned the responsibility of professional translators for the final translated text. Participants generally viewed GenAI as a supporting technology rather than a replacement for professional judgement. As one respondent explained: “The concept of translation competence needs to be redefined in the age of AI: the competent translator is no longer simply one who can translate well, but one who knows how and when to use AI tools

consciously, evaluates their output critically, and takes the final decision on the basis of his linguistic, cultural, and ethical judgment" (P14). Some participants also expressed concern that excessive reliance on AI could weaken independent translation skills. These responses highlight the perceived importance of critical evaluation and professional responsibility in human–AI collaboration.

Automated Metric Analysis: BLEU and chrF

Overview and Rationale

BLEU (Papineni et al., 2002) and chrF (Popović, 2015) were used to measure the similarity between GPT-4o-generated translations and professionally published human translations across 45 parallel texts. BLEU measures word-level n-gram overlap with a reference translation, while chrF measures character-level n-gram similarity, making it useful for examining morphologically complex languages such as Arabic. Both measures have been employed in recent Arabic machine translation research (Jon et al., 2026; Elchafei & Fashwan, 2026). However, these metrics measure similarity to a reference translation rather than translation quality in its full sense. Consequently, the results are interpreted as indicators of textual similarity rather than independent measures of translation accuracy or adequacy.

Corpus-Level and Register-Level Results

Across the 45 texts, GPT-4o achieved a mean document-level BLEU score of 14.87 (SD = 6.26) and a mean chrF score of 59.27 (SD = 6.29). Considerable variation was observed across registers and individual texts, with BLEU scores ranging from 5.75 to 29.34. As automated evaluation scores are sensitive to reference translations, tokenisation, and evaluation procedures, the findings are interpreted within the present corpus rather than against predefined quality thresholds (Post, 2018; Reiter, 2018). Table 3 presents the results by register.

Table 3

Mean Document-Level BLEU and chrF Scores by Register (N = 45 Texts)

Domain	n	BLEU Mean (SD)	chrF Mean (SD)	Range (BLEU)
Legal	9	23.74 (4.71)	66.48 (5.69)	15.41–29.34
Literary	12	11.95 (1.74)	54.86 (2.89)	8.74–14.44
Media	12	12.86 (4.59)	59.23 (5.64)	5.75–22.40
Scientific	12	13.13 (5.94)	58.34 (5.46)	6.39–28.30
Overall	45	14.87 (6.26)	59.27 (6.29)	5.75–29.34

Note. Values represent means and sample standard deviations of scores calculated separately for each text. They should not be confused with aggregated corpus-level BLEU and chrF scores.

Legal texts recorded the highest mean scores on both metrics (BLEU: 23.74; chrF: 66.48). This may partly reflect the register-specific linguistic conventions and terminology characteristic of legal writing (Cao, 2007; Tiersma, 1999). However, variation was observed within the legal register, with BLEU scores ranging from 15.41 to 29.34. This suggests that the degree of similarity between AI-generated and human translations may vary even within the same register, depending on the linguistic and textual characteristics of individual documents.

Literary texts recorded the lowest mean BLEU and chrF scores (11.95 and 54.86, respectively), with relatively limited variation between individual texts. These findings indicate that GPT-4o-generated literary translations exhibited lower overlap with their human reference translations than those in the other registers. This pattern may partly reflect differences in lexical and stylistic choices, which are particularly relevant in literary translation. However, lower automated scores do not necessarily indicate lower translation quality, since acceptable

stylistic variation may also reduce similarity to a particular reference translation (Wang et al., 2021).

Media texts produced a mean BLEU score of 12.86 (SD = 4.59) and a mean chrF score of 59.23 (SD = 5.64). Although their BLEU scores were relatively close to those of literary texts, the higher chrF scores indicate greater character-level similarity to the human reference translations. These results suggest differences in the extent of lexical and character-level overlap across the two registers, although the underlying linguistic factors cannot be established from the automated scores alone.

Scientific texts recorded a mean BLEU score of 13.13 (SD = 5.94) and a mean chrF score of 58.34 (SD = 5.46). The relatively high variation in BLEU scores, ranging from 6.39 to 28.30, suggests considerable differences between individual scientific texts in their degree of similarity to the human reference translations. Such variation may be associated with differences in terminology, subject matter, or textual characteristics, although further qualitative examination would be required to establish the factors involved.

Taken together, the automated evaluation results indicate that similarity between GenAI-generated and human translations varies across registers and individual texts. Legal translation showed the highest average similarity, whereas literary translation exhibited the lowest. These findings provide a quantitative basis for further examining lexical and contextual differences between the two translation conditions.

Corpus-Based Lexical Analysis

Word List Comparison: Lexical Diversity

Word List comparison between the HT_Corpus (213,111 tokens) and AI_Corpus (199,958 tokens) revealed differences in lexical diversity. The human translation corpus contained 12,196 distinct word types, compared with 9,590 in the AI corpus. Type-Token Ratio

(TTR) values were 5.72% and 4.80%, respectively, representing a difference of 0.92 percentage points.

To account more effectively for differences in text length, Moving-Average Type-Token Ratio (MATTR) was also calculated using a moving window of 100 words within individual files (Covington & McFall, 2010). The human translation corpus recorded a MATTR of 84.79%, compared with 83.71% for the AI corpus, a difference of 1.08 percentage points. Table 4 summarises these results.

Table 4

Lexical Diversity Measures: HT_Corpus and AI_Corpus

Corpus	Types	Tokens	TTR (%)	MATTR (%)	TTR Diff.	MATTR Diff.
HT_Corpus	12,196	213,111	5.72	84.79	+0.92 pp	+1.08 pp
AI_Corpus	9,590	199,958	4.80	83.71	—	—

Note. MATTR was calculated within individual files using a moving window of 100 words (Covington & McFall, 2010). TTR = types ÷ tokens × 100; pp = percentage points.

Both measures indicate greater lexical diversity in the human translation corpus. Although raw TTR is sensitive to corpus size, the corresponding MATTR results provide further support for this difference. These findings suggest that professional human translations in the present corpus exhibit somewhat greater lexical variation than GPT-4o-generated translations.

Keyword Analysis: Distinctive Lexical Patterns

Bidirectional Keyword Analysis was conducted using Sketch Engine (Kilgarriff et al., 2014), with each translation corpus serving as the reference corpus for the other. Keywords were ranked according to their keyness scores to identify words occurring at higher relative frequencies in one corpus than in the other. The selected keywords were subsequently examined in context and classified according to their linguistic functions.

Keywords Distinctive of Human Translation (HT_Corpus)

Table 5

Top Keywords Distinguishing HT_Corpus from AI_Corpus (Single Words)

HT Keyword	Gloss	Freq. (HT)	Keyness Score	Category
ثمة	there is / there are	49	230.93	Discourse marker
صار	became / came to be	45	212.16	Stylistic verb
اضطلع	undertook / assumed	29	136.08	Formal verb
نفع	benefit / avail	20	93.85	Semantic precision
توجب	necessitated / required	19	89.16	Modal/legal verb
لعل	perhaps / it may be	18	84.46	Epistemic modal
إنما	rather / only	15	70.39	Restrictive particle
أفضى	led to / resulted in	13	61.00	Causative verb
استدرك	corrected / rectified	12	56.31	Discourse-related verb
بغية	with the aim of	10	46.92	Purpose marker

The keywords distinctive of human translation suggest several lexical and stylistic patterns. First, the relatively frequent use of expressions such as ثمة *thamma* (‘there is/there are’), إنما *innamā* (‘rather/only’), and لعل *laʿalla* (‘perhaps/it may be’) indicates differences in the use of presentational, restrictive, and modal expressions. These items contribute to the organisation of information and the expression of textual relations in Arabic.

Second, verbs such as أفضى *afḍā* (‘led to/resulted in’), اضطلع *iḍṭalaʿa* (‘undertook/assumed’), and استدرك *istadrak* (‘corrected/rectified’) occur at higher relative frequencies in the human translation corpus. These verbs are associated with formal and register-sensitive uses of Arabic and may reflect differences in lexical preferences between human and AI-generated translations.

Third, expressions such as توجب *tawajjab* (‘necessitated/required’) and بغية *bughyat* (‘with the aim of’) suggest further differences in the selection of modal and purposive expressions. Taken together, these patterns provide examples of lexical choices that may be relevant to Cultural-Communicative Competence and Terminological Precision Competence, particularly where register and contextual appropriateness are important.

Keywords Distinctive of AI Translation (AI_Corpus)

Table 6

Top Keywords Distinguishing AI_Corpus from HT_Corpus (Single Words)

Rank	AI Keyword	Freq. (AI)	Keyness Score	Category
1	فاذرينغاي (Fatheringay)	87	436.09	Transliterated proper noun
2	اميسوم (AMISOM)	53	266.06	Transliterated acronym
3	جينسن (Jensen)	44	221.05	Transliterated proper noun
4	متأثر (affected)	14	71.01	General adjective
5	يائس (desperate)	7	36.01	General adjective
6	ملخص (summary)	7	36.01	Meta-textual term
7	تمثيلي (representative)	7	36.01	General noun
8	ديكتاتوري (dictatorial)	5	26.00	General adjective
9	ترفيه (entertainment)	5	26.00	General adjective
10	تنصيب (installation)	5	26.00	General noun

The keywords distinctive of AI translation suggest two main patterns. The first concerns the treatment of foreign proper nouns and acronyms. Items such as فاذرينغاي *Fādhiringhāy* (‘Fatheringay’), اميسوم AMISOM, and جينسن *Jīnsin* (‘Jensen’) recorded the highest keyness scores, indicating differences in their relative use between the two translation corpora. Examination of corresponding translations suggests differences in transliteration and the treatment of acronyms. For example, the human translation rendered AMISOM as بعثة الاتحاد الأفريقي في الصومال *Baʿthat al-Ittiḥād al-Afrīqī fī al-Ṣūmāl* (‘the African Union Mission in Somalia’), whereas the AI translation retained a transliterated form. Such differences may have implications for cultural and terminological appropriateness, depending on the communicative context.

The second pattern concerns the relatively frequent occurrence of general lexical items in the AI corpus. Words such as متأثر *mutaʾaththir* (‘affected’), يائس *yāʾis* (‘desperate’), and تمثيلي *tamthīlī* (‘representative’) appeared among the most distinctive keywords. Their distribution suggests differences in lexical selection between AI-generated and human translations, although their contextual appropriateness cannot be determined from frequency

comparisons alone. The qualitative examination of these items indicates areas where further attention to register, context, and lexical specificity may be required.

These findings suggest that the evaluation of AI-generated translations should consider not only grammatical correctness and fluency, but also the appropriateness of lexical choices, proper-noun treatment, and terminology in different registers.

Implications for the HACTC Framework

The corpus findings provide evidence of differences between human and AI-generated translations in lexical diversity, the use of discourse-related expressions, and the treatment of proper nouns and terminology. These differences are relevant to several competency areas considered in the proposed HACTC Framework.

The lexical and contextual patterns identified in the human translation corpus highlight the importance of Cultural-Communicative Competence and Terminological Precision Competence. Differences in the AI corpus indicate potential areas requiring post-editing and critical evaluation, particularly where fluent translations may contain lexical or contextual choices that require further examination.

Taken together, these findings provide support for including corpus-based analytical skills in translator training, particularly as part of AI Digital Competence. These findings are considered further in the proposed HACTC Framework presented below.

The Human–AI Collaborative Translation Competence (HACTC) Framework

Theoretical Foundations and Design Rationale

Based on the findings of the present study, the Human–AI Collaborative Translation Competence (HACTC) Framework is proposed to account for the knowledge and skills required in GenAI-mediated translation practice. The framework draws on the critical review

of PACTE and EMT, the survey findings concerning the perceived importance of AI-related competencies and their coverage in translator training, and the corpus analysis of differences between human and AI-generated translations. These sources provide complementary evidence for identifying competencies that require greater attention in contemporary translation practice.

The proposed framework builds on PACTE and EMT rather than replacing them. Both models recognise the importance of linguistic, extra-linguistic, strategic, and technological knowledge in translation. However, as discussed in the literature review, the increasing use of GenAI introduces requirements that are not explicitly or fully addressed by these frameworks, particularly the formulation of instructions for generative systems, the critical evaluation of their output, and the exercise of professional judgement in human–AI translation workflows.

Three considerations inform the development of HACTC. The first is the relationship between the proposed competencies and the findings of the present study. The framework draws on both participants' reported needs and the linguistic differences observed in the parallel corpus, while recognising that these sources provide different forms of evidence. The second is the relevance of the competencies to translator training. Each component is defined in terms of knowledge and skills that can inform learning outcomes and practical training activities. The third is compatibility with existing competence frameworks. Rather than introducing an entirely separate set of professional requirements, HACTC brings together established competencies and additional skills associated with GenAI, with particular attention to the relationships between them.

Framework Structure: Six Competencies and a Cross-Cutting Ethical Component

The HACTC Framework consists of six main competencies organised into two functional groups, together with Ethical Competence as a seventh, cross-cutting component. The first group, AI Operational Competencies, concerns the knowledge and skills required to

use generative AI tools in translation workflows. It comprises AI Digital Competence and Prompt Engineering Competence. The second group, Critical-Evaluative Competencies, concerns the evaluation, revision, and contextual appropriateness of AI-generated translations. It comprises Post-editing Competence, Critical Evaluation Competence, Cultural-Communicative Competence, and Terminological Precision Competence. Ethical Competence extends across both groups because questions of confidentiality, disclosure, accountability, and professional responsibility arise throughout the translation process, from the selection and use of AI tools to the evaluation and delivery of translated texts.

Table 7 summarises the competencies, their proposed functions, and the evidence informing their inclusion in the framework.

Table 7

The HACTC Framework: Competencies, Functions, and Supporting Evidence

Competency	Group	Core function	Survey importance (M/5)	Training gap	Supporting evidence
AI Digital Competence	AI Operational	Selecting and using GenAI tools with an understanding of their capabilities and limitations	3.77	0.71	AI Literacy survey ratings; variation in reference similarity across texts
Prompt Engineering Competence	AI Operational	Formulating and refining instructions for AI-assisted translation tasks	3.87	1.00	Prompt Engineering survey ratings and qualitative responses
Post-editing Competence	Critical-Evaluative	Reviewing and revising AI-generated translations in accordance with textual and professional requirements	3.87	0.90	Survey ratings; differences between AI and human translations
Critical Evaluation Competence	Critical-Evaluative	Assessing the accuracy, appropriateness, and limitations of AI-generated output	3.91	1.05	Highest survey importance rating; automated and lexical analysis
Cultural-Communicative Competence	Critical-Evaluative	Evaluating cultural references, contextual meanings, discourse choices, and register appropriateness	3.84	1.06	Cultural Awareness survey ratings; lexical and contextual differences
Terminological Precision Competence	Critical-Evaluative	Assessing and selecting contextually appropriate terminology, particularly in specialised translation	N/A*	N/A*	Corpus-based lexical analysis; theoretical review
Ethical Competence (cross-cutting)	Both groups	Addressing confidentiality, disclosure, accountability, and responsible AI use	3.77	0.78	Ethical Competence survey ratings and qualitative responses

The survey assessed six preliminary competency areas rather than the final HACTC components in their present form. AI Literacy and Cultural Awareness inform the broader AI Digital and Cultural-Communicative competencies, respectively. Terminological Precision Competence was not measured as a separate survey category; therefore, no independent importance rating or training gap is reported for it. The supporting evidence indicates the basis for proposing each component rather than independent empirical validation of the framework.

Component Definitions and Empirical Grounding

AI Digital Competence

AI Digital Competence refers to the knowledge and skills required to select, use, and evaluate generative AI tools for translation purposes. It includes an understanding of how these tools operate, their potential applications, and their limitations in different translation contexts. Translators also need to recognise when AI assistance is appropriate and when the linguistic, cultural, or professional requirements of a task call for greater human involvement.

This competency builds on the technological knowledge already recognised in existing translation competence frameworks. Its inclusion in HACTC is supported by the survey findings for AI Literacy (M = 3.77; Gap = 0.71) and by the corpus results, which showed differences in similarity to human reference translations across registers and individual texts. Such variation does not establish the technical causes of particular translation choices, but it highlights the importance of understanding that AI output may differ across tasks and textual contexts.

Prompt Engineering Competence

Prompt Engineering Competence is defined as the ability to formulate, refine, and evaluate instructions given to generative AI systems to support translation tasks. It includes

specifying the purpose of the translation, intended audience, register, terminology, and other relevant requirements, as well as revising instructions when the initial output does not meet the needs of the task.

Participants considered this competency important (M = 3.87), while reporting comparatively limited training opportunities (Gap = 1.00). Their qualitative responses also identified prompt formulation as an area requiring greater attention in translator education. Although the present corpus analysis employed a defined translation prompt, it did not compare different prompting strategies. The inclusion of Prompt Engineering Competence is therefore supported primarily by the survey findings and the practical requirements of working with generative AI systems, rather than by a measured effect of prompt design on translation quality.

Post-editing Competence

Post-editing Competence refers to the ability to review and revise AI-generated translations in accordance with the linguistic, functional, and professional requirements of a translation task. It involves identifying problems in AI output, determining the extent of revision required, and making appropriate corrections while preserving the meaning and communicative purpose of the source text.

The survey findings indicate that participants regarded post-editing as an important competency (M = 3.87), with a perceived training gap of 0.90. The corpus analysis also provides examples of lexical and textual differences that may require examination during post-editing. In addition, the variation in BLEU and chrF scores across individual texts suggests that the degree of similarity to human reference translations is not uniform, even within the same register. These findings support the relevance of text-sensitive evaluation and revision, although the present study did not directly measure post-editing effort or time.

Critical Evaluation Competence

Critical Evaluation Competence is defined as the ability to assess AI-generated translations against linguistic, contextual, and professional criteria. It involves identifying potential inaccuracies, evaluating alternative translations, and distinguishing between errors and acceptable variations in expression. It also requires awareness of the limitations of automated evaluation measures and the need for informed human judgement.

Critical Evaluation received the highest importance rating in the survey (M = 3.91), with a training gap of 1.05. The corpus findings provide further support for its relevance. The differences observed in BLEU and chrF scores across registers demonstrate that similarity to a human reference varies between texts, while the lexical analysis identifies differences that require contextual interpretation. Neither type of analysis independently establishes whether a particular translation is acceptable. This reinforces the importance of translators' ability to evaluate AI-generated output using linguistic knowledge, contextual understanding, and professional judgement rather than relying exclusively on automated scores or apparent fluency.

Cultural-Communicative Competence

Cultural-Communicative Competence refers to the ability to evaluate and adapt AI-generated translations in relation to the cultural, pragmatic, and communicative requirements of the target context. It includes the treatment of cultural references, idiomatic expressions, implicit meanings, register, discourse organisation, and proper nouns when these affect the interpretation or accessibility of a translated text.

This competency draws on the existing cultural and extra-linguistic dimensions of translator competence, while giving greater attention to their application in AI-mediated translation. Cultural Awareness recorded the largest perceived training gap in the survey (1.06), indicating that participants considered further training in this area particularly important.

Qualitative responses also emphasised difficulties involving cultural references, local conventions, and contextual meaning.

The lexical analysis offers additional examples relevant to this competency. The human translation corpus showed higher relative frequencies of expressions such as ثمة*thamma*, إنما *innamā*, and لعل*laʿalla*, which perform different presentational, restrictive, and modal functions in Arabic. Differences were also observed in the treatment of proper nouns and acronyms, including AMISOM, for which the human translation used an explanatory Arabic form while the AI translation employed transliteration. These patterns do not establish that every AI choice was inappropriate, but they identify areas where cultural and communicative considerations may affect translation decisions.

Terminological Precision Competence

Terminological Precision Competence refers to the ability to assess and select terminology appropriate to the subject matter, register, and communicative purpose of a translated text. It includes recognising specialised meanings, evaluating competing equivalents, and maintaining terminological consistency where required. In AI-mediated translation, this competency also involves examining fluent but potentially imprecise lexical choices that may not adequately reflect the terminology of a particular field.

Unlike the other components, Terminological Precision Competence was not assessed as a separate category in the survey. Its inclusion is informed by the theoretical review and the corpus-based analysis of lexical differences between human and AI-generated translations. The distinctive keyword profiles, including differences involving formal verbs and specialised expressions, suggest the importance of evaluating lexical choices in context. The variation observed in scientific and legal texts provides further reason to consider terminology in the evaluation of AI-generated translations, although automated similarity scores cannot determine terminological accuracy on their own.

The competency is therefore proposed as a distinct area of attention within HACTC, while remaining closely related to Post-editing and Critical Evaluation Competence.

Ethical Competence (Cross-Cutting)

Ethical Competence refers to the ability to make responsible decisions about the use of GenAI throughout the translation process. It includes understanding obligations relating to confidentiality, data protection, disclosure of AI assistance, professional accountability, and the appropriate use of client or institutional information. It also involves maintaining independent professional judgement and recognising the possible effects of excessive reliance on AI tools.

Survey participants considered Ethical Competence important (M = 3.77), with a reported training gap of 0.78. The qualitative responses further raised concerns relating to confidentiality, professional responsibility, and the maintenance of translators' own skills. These considerations are not confined to a single stage of translation. They apply when selecting AI tools, preparing source texts for processing, formulating prompts, evaluating output, and delivering the final translation.

For this reason, Ethical Competence is treated as a cross-cutting component that informs both the AI Operational and Critical-Evaluative groups rather than as a separate stage of the translation process.

Discussion

Overview: Evidence from the Three Components

The findings presented in the Results section, considered alongside the critical review of existing competence frameworks, provide complementary evidence of the changing requirements of translator competence in GenAI-mediated translation practice. The survey findings indicate that participants recognise the increasing importance of evaluating and

supervising AI-generated translations, while the corpus analysis identifies lexical and textual differences between professional human translations and GPT-4o-generated translations. Although these components address different aspects of translation practice, their findings provide support for reconsidering the relative importance of existing competencies and identifying additional requirements associated with human–AI collaboration. This section discusses the findings and the proposed HACTC Framework in relation to the four research questions and existing literature, with particular attention to their implications for translator competence, training, and professional practice.

RQ1: How Is GenAI Integration Reshaping the Professional Role of Translators?

The survey findings suggest that the increasing use of GenAI is contributing to changes in the professional role of translators, particularly in relation to the evaluation, revision, and supervision of AI-generated output. Most respondents (89.8%) reported using GenAI tools either daily or occasionally, indicating that these technologies were widely used within the surveyed sample. The qualitative responses further suggest that participants associate professional competence not only with the ability to produce translated texts, but also with the ability to evaluate AI-generated translations and make informed decisions about their appropriateness. As one participant observed, the competent translator is increasingly expected to know when and how to use AI tools, critically evaluate their output, and exercise professional judgement over the final translation (P14).

These findings are in line with earlier research on the changing organisation of translation practice through technological and collaborative developments (Jiménez-Crespo, 2017). However, the present study extends this discussion by considering the particular competencies associated with GenAI-mediated translation in the Arabic–English context. The corpus findings provide further evidence relevant to these requirements, particularly the

differences observed in lexical diversity, register-sensitive expressions, and the treatment of proper nouns and acronyms. Although these textual differences do not directly demonstrate changes in professional roles, they identify areas where human evaluation and intervention may be required. The findings therefore suggest that evaluative and supervisory responsibilities are becoming increasingly important alongside established translation production skills. This finding is consistent with Prieto Ramos (2024), who reconsidered translator competence in relation to developments in translation technologies, particularly in legal and institutional settings. While that study approached the issue from a specific professional context, the present findings suggest that the growing importance of evaluative responsibilities is also recognised by a broader group of translators, academics, and postgraduate students. The findings therefore support the need to reconsider how established professional competencies are applied in GenAI-mediated translation practice.

RQ2: What Specific Competencies Are Required for Effective Human–AI Collaboration in Translation?

The findings provide support for several competency areas relevant to human–AI collaborative translation, with the different components of the study contributing evidence from different perspectives. Critical Evaluation received the highest importance rating in the survey (M = 3.91), while Cultural Awareness recorded the largest perceived training gap (1.06). These findings suggest that participants recognise the importance of evaluating AI-generated translations beyond grammatical correctness and linguistic fluency, particularly where cultural and contextual considerations are involved.

The corpus analysis provides further support for the importance of these competencies. The automated evaluation showed considerable variation in similarity to human reference translations across registers, with legal texts recording the highest mean scores and literary texts the lowest. Although BLEU and chrF cannot independently determine translation quality

or identify specific translation errors, these differences indicate that the relationship between human and AI-generated translations varies across textual contexts. This variation highlights the importance of evaluating translations in relation to their particular linguistic and communicative requirements rather than relying on general assessments of AI performance.

The lexical analysis provides more specific evidence of differences in translation choices. Human translations exhibited somewhat greater lexical diversity, as indicated by both TTR and MATTR, while Keyword Analysis identified differences in the use of discourse-related expressions, formal vocabulary, proper nouns, and acronyms. The treatment of items such as AMISOM illustrates how differences in lexical choices may affect the accessibility and appropriateness of translated texts for Arabic readers. These findings are particularly relevant to Cultural-Communicative Competence and Terminological Precision Competence, as they indicate areas where contextual interpretation and informed lexical choices remain important. These findings are broadly consistent with previous studies highlighting the limitations of GenAI-generated translations in culturally and contextually sensitive tasks. Khoshafah (2023) identified difficulties in preserving contextual, historical, and cultural nuances, while Sahari et al. (2024) reported advantages for human translators in conveying figurative and idiomatic meanings. Although the present study examined different textual features, its findings similarly highlight the importance of contextual and cultural judgement. However, the observed lexical differences do not necessarily indicate translation errors, as their appropriateness depends on the source text, register, and communicative purpose.

Post-editing Competence and Critical Evaluation Competence are also relevant to identifying and addressing such differences in AI-generated translations. However, the corpus findings do not directly measure post-editing effort or the cognitive processes involved in translation. Rather, they identify textual patterns that translators may need to evaluate and revise. AI Digital Competence and Prompt Engineering Competence receive further support

from the survey responses, particularly participants' recommendations for greater practical training in AI tools and prompt formulation. Taken together, these findings suggest that effective human–AI collaboration requires both established translation competencies and additional knowledge and skills associated with generative technologies.

RQ3: To What Extent Do Existing Competence Frameworks Adequately Account for the Demands of Human–AI Collaborative Practice?

The critical review indicates that PACTE and EMT provide important theoretical and professional foundations for understanding translator competence, although their treatment of the requirements associated with GenAI-mediated translation remains limited. PACTE's strategic sub-competence, which involves planning, monitoring, and evaluating the translation process, provides a relevant basis for understanding the evaluative and supervisory responsibilities identified in the present study. However, the model does not explicitly specify the knowledge and skills required for working with generative AI systems, particularly prompt formulation and the critical evaluation of AI-generated output.

Similarly, the EMT framework recognises the importance of translation technologies, including machine translation and post-editing, and therefore provides an existing basis for developing AI-related competencies. Nevertheless, the integration of GenAI into translation practice raises questions that extend beyond the instrumental use of translation technologies, particularly concerning the relationships between technological knowledge, strategic competence, cultural interpretation, and professional responsibility.

The findings of the present study suggest that these competencies should not be considered in isolation. Survey participants emphasised the importance of critical evaluation and cultural awareness, while the corpus findings identified lexical and contextual differences relevant to these areas. This provides support for extending existing competence frameworks to account more explicitly for the interaction between technological, linguistic, cultural, and

evaluative competencies in human–AI collaborative translation. The proposed HACTC Framework builds on these existing foundations by identifying additional competency requirements and reconsidering how the different components of translator competence relate to one another. This proposed extension can also be considered in relation to previous research on established competence frameworks. Chodkiewicz (2012) examined how translators and translation students perceived the relevance of EMT competencies, while Krause (2024) considered the applicability of the 2022 EMT framework in response to changing professional requirements. The present study builds on this line of research by drawing on participants' perceptions and corpus-based evidence to identify competencies requiring more explicit attention in GenAI-mediated translation. Rather than challenging the foundations of existing models, these findings suggest the value of specifying how technological, evaluative, and contextual competencies interact in human–AI collaborative workflows.

Several HACTC components are closely related to existing competencies. For example, Critical Evaluation and Post-editing Competence draw on the monitoring, problem-solving, and decision-making functions associated with PACTE's strategic sub-competence. Similarly, AI Digital Competence builds on the technological knowledge recognised by EMT, while Cultural-Communicative and Terminological Precision Competence extend established linguistic and extra-linguistic requirements into AI-mediated contexts.

More specifically, HACTC identifies and organises competencies according to their functions in human–AI translation workflows. Prompt Engineering makes explicit a requirement associated with interacting with generative systems, while the Critical-Evaluative group brings together the skills needed to assess and revise AI-generated translations. The cross-cutting position of Ethical Competence further emphasises the relationship between technological decisions and professional responsibility.

RQ4: What Are the Implications for Translator Training and Professional Benchmarking?

The survey findings have implications for translator training, particularly in relation to the perceived differences between the importance of emerging competencies and the availability of relevant training. Across all six competency areas assessed, importance ratings exceeded training availability ratings, with gaps ranging from 0.71 to 1.06. Although these findings reflect participants' perceptions rather than a direct evaluation of existing curricula, they suggest that greater attention may be required to the integration of GenAI-related competencies into translator education.

The qualitative findings provide further indications of the areas that participants considered particularly important. These included practical training in prompt engineering, systematic post-editing, critical evaluation of AI-generated translations, and the ethical use of GenAI. Participants also emphasised that such training should complement rather than replace established linguistic, cultural, and translation skills. As one respondent stated, translator training should include post-editing and critical evaluation while continuing to develop the foundational skills required to distinguish appropriate from inappropriate translations (P58).

These findings suggest that translator education may benefit from integrating GenAI-related knowledge and skills across existing training components rather than treating them exclusively as separate technological skills. The corpus analysis also indicates the potential value of corpus-based analytical methods in developing translators' ability to identify lexical and contextual differences between human and AI-generated translations. In addition, professional benchmarking may need to give greater consideration to the ability to evaluate, revise, and take responsibility for AI-assisted translation output. The HACTC Framework provides a basis for organising these competency requirements and considering their implications for translator training and professional standards.

The framework also offers a basis for developing learning outcomes and assessment activities in translator education. AI Operational Competencies may be addressed through practical tasks involving tool selection and prompt formulation, while Critical-Evaluative Competencies may be assessed through comparative translation analysis, post-editing exercises, and contextual evaluation of AI-generated output. Ethical Competence can be incorporated throughout these activities by requiring students to justify their decisions about AI use, confidentiality, and responsibility for the final translation.

Conclusion and Recommendations

Summary of Principal Findings

This study investigated the implications of generative AI for translator competence through a convergent mixed-methods design combining a critical review of existing competence frameworks, a survey of 98 respondents, and a comparative corpus analysis of professional human and GPT-4o-generated Arabic translations. The findings suggest that the increasing use of GenAI places greater emphasis on translators' evaluative and supervisory responsibilities alongside their established role in text production. Survey participants considered critical evaluation and cultural awareness particularly important, while reporting gaps between the perceived importance of AI-related competencies and the availability of relevant training. The corpus analysis identified differences between human and AI-generated translations in lexical diversity, register-sensitive expressions, and the treatment of proper nouns and acronyms. These findings indicate areas where linguistic, cultural, and contextual judgement remains important in AI-mediated translation practice.

The critical review further indicated that PACTE and EMT provide relevant foundations for understanding translator competence, although some requirements associated with GenAI-mediated translation are not explicitly addressed in these frameworks. Drawing on

the theoretical review and empirical findings, the study proposed the Human–AI Collaborative Translation Competence (HACTC) Framework, consisting of six competencies organised into two functional groups, with Ethical Competence as a seventh, cross-cutting component.

Theoretical Contributions

The main contribution of this study is the proposed HACTC Framework, which brings together established translation competencies and additional requirements associated with human–AI collaboration. Rather than replacing PACTE or EMT, the framework extends their existing foundations by placing greater emphasis on the interaction between technological knowledge, critical evaluation, linguistic and cultural judgement, and professional responsibility. It also defines the proposed competencies in terms that may inform translator training and professional assessment.

A further contribution lies in the use of complementary methods to examine translator competence. The survey provides evidence of perceived competency requirements and training needs, while the corpus analysis identifies textual differences relevant to translation evaluation. Considered alongside the theoretical review, these findings provide a basis for proposing a competence framework that addresses several dimensions of GenAI-mediated translation practice.

Practical Recommendations

The findings suggest that translator training programmes should integrate GenAI-related competencies into existing curricula while maintaining a strong foundation in linguistic, cultural, and translation skills. Training in AI tool use and prompt formulation should be accompanied by practical exercises in post-editing, critical evaluation, contextual interpretation, and terminological decision-making. Such activities may draw on authentic AI-generated translations and combine automated measures with qualitative assessment. Ethical

considerations, particularly confidentiality, disclosure, and professional accountability, should be addressed throughout the training process.

The proposed framework may also inform the development of professional assessment activities that examine translators' ability to evaluate and revise AI-generated output. For practising translators, the findings highlight the importance of maintaining independent professional judgement and developing the knowledge and skills needed to use GenAI effectively and responsibly.

Limitations and Directions for Future Research

Several limitations should be considered when interpreting the findings. The survey involved 98 respondents drawn from a particular academic and professional context, which limits the generalisability of the results. The corpus analysis was restricted to Arabic translations generated by GPT-4o under a single prompting configuration and compared with professionally published human translations. BLEU and chrF measured similarity to the human reference translations rather than translation quality independently, while TTR, MATTR, and keyword analysis identified lexical patterns without establishing their contextual appropriateness in every instance. The study also did not directly measure post-editing effort or validate the proposed competencies through performance-based assessment.

Future research could examine additional language pairs, GenAI systems, and prompting strategies, alongside human evaluation of translation quality and post-editing effort. Further studies could also investigate the proposed HACTC components through the development and validation of competency assessment instruments and the evaluation of training programmes designed around the framework. Such research would help establish the applicability of HACTC across different translation settings and clarify its potential contribution to translator education and professional practice.

Funding Statement

This research received no external funding.

Conflict of Interest Statement

The author declares no conflict of interest.

Data Availability Statement

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

AI Disclosure Statement

The author used Claude (Anthropic) for language editing and structural feedback during manuscript preparation. The AI tool did not independently generate, analyse, or interpret any research data, findings, or theoretical contributions reported in this study. All content, arguments, interpretations, scholarly judgements, data collection, data analysis, citations, and bibliographic references were conducted, reviewed, verified, and approved by the author.

Bio

Dr. Ibrahim Alasmri is an Assistant Professor of Translation Studies and Vice Dean of Educational Affairs and Development at the College of Languages and Translation, King Khalid University, Saudi Arabia. He holds a PhD in Translation Studies from Macquarie University, Australia. His scholarly work sits at the intersection of corpus-based translation studies, Systemic Functional Linguistics, and language variation and change, with a growing research agenda addressing the implications of Generative Artificial Intelligence for translation competence, training, and professional practice.

References

Alasmri, I. (2024). Cross-linguistic influence of English on conjunction system in translated Arabic: A multifactorial corpus-based study. *Perspectives, 34*(5), 934–956. https://doi.org/10.1080/0907676X.2024.2427682

Al-Harthi, M., Alsaif, A., Al-Nafjan, E., Alshihri, F., & Saleh, M. (2024). Saudi Learner Translation Corpus: The design and compilation of an English–Arabic learner translation corpus. *PLOS ONE, 19*(10), Article e0303729. https://doi.org/10.1371/journal.pone.0303729

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*(3), 837–849. https://doi.org/10.30564/fls.v6i3.6799

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cao, D. (2007). *Translating law*. Multilingual Matters. https://doi.org/10.21832/9781853599552

Chodkiewicz, M. (2012). The EMT framework of reference for competences applied to translation: Perceptions by professional and student translators. *The Journal of Specialised Translation*, (17), 37–54. https://doi.org/10.26034/cm.jostrans.2012.458

Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics, 17*(2), 94–100. https://doi.org/10.1080/09296171003643098

Elchafei, P., & Fashwan, A. (2026). Arabic ChartSumm: An English-to-Arabic benchmark for metadata-to-text summarization. In *Proceedings of the Fifteenth Language Resources and Evaluation Conference* (pp. 10447–10456). European Language Resources Association. https://doi.org/10.63317/5pxwxs9vtum4

EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. European Commission.

EMT Expert Group. (2022). *EMT competence framework 2022*. European Commission.

Habash, N., Marzouk, R., Khairallah, C., & Khalifa, S. (2022). Morphotactic modeling in an open-source multi-dialectal Arabic morphological analyzer and generator. In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology* (pp. 92–102). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.sigmorphon-1.10

Hazaea, A. N., & Qassem, M. (2024). Translation competence in translator training programs at Saudi universities: Empirical study. *Open Education Studies, 6*(1), Article 20240029. https://doi.org/10.1515/edu-2024-0029

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

Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties* (Rev. ed.). Georgetown University Press.

Jiao, W., Wang, W., Huang, J.-t., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator? Yes with GPT-4 as the engine*. arXiv. https://doi.org/10.48550/arXiv.2301.08745

Jiménez-Crespo, M. A. (2017). *Crowdsourcing and online collaborative translations: Expanding the limits of translation studies*. John Benjamins. https://doi.org/10.1075/btl.131

Jon, J., Bondok, R., & Bojar, O. (2026). Current state of LLMs for Arabic dialectal machine translation. In *Proceedings of the 2nd Workshop on NLP for Languages Using Arabic Script* (pp. 329–363). Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.abjadnlp-1.41

Karpinska, M., & Iyyer, M. (2023). Large language models effectively leverage document-level context for literary translation, but critical errors persist. In *Proceedings of the Eighth Conference on Machine Translation* (pp. 419–451). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.wmt-1.41

Khoshafah, F. (2023). *ChatGPT for Arabic-English translation: Evaluating the accuracy*. Research Square. https://doi.org/10.21203/rs.3.rs-2814154/v2

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*(1), 7–36. https://doi.org/10.1007/s40607-014-0009-9

Kiraly, D. C. (2000). *A social constructivist approach to translator education: Empowerment from theory to practice*. St. Jerome Publishing.

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), 131–148. https://doi.org/10.26034/cm.jostrans.2016.303

Krause, A. (2024). The 2022 EMT competence framework: An outlook on its applicability in a rapidly evolving translation landscape. *Alkalmazott Nyelvtudomány, 2024*(3), 1–14. https://doi.org/10.18460/ANY.K.2024.3.001

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

Läubli, S., Sennrich, R., & Volk, M. (2018). Has machine translation achieved human parity? A case for document-level evaluation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 4791–4796). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-1512

O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces, 1*, 101–122. https://doi.org/10.1075/ts.1.05obr

OpenAI. (2024). *GPT-4o system card*. arXiv. https://doi.org/10.48550/arXiv.2410.21276

PACTE Group. (2003). Building a translation competence model. In F. Alves (Ed.), *Triangulating translation: Perspectives in process oriented research* (pp. 43–66). John Benjamins. https://doi.org/10.1075/btl.45.06pac

PACTE Group. (2005). Investigating translation competence: Conceptual and methodological issues. *Meta, 50*(2), 609–619. https://doi.org/10.7202/011004ar

PACTE Group. (2020). Translation competence acquisition. Design and results of the PACTE group's experimental research. *The Interpreter and Translator Trainer, 14*(2), 95–233. https://doi.org/10.1080/1750399X.2020.1732601

Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318). Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135

Popović, M. (2015). chrF: Character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 392–395). Association for Computational Linguistics. https://doi.org/10.18653/v1/W15-3049

Post, M. (2018). A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers* (pp. 186–191). Association for Computational Linguistics. https://doi.org/10.18653/v1/W18-6319

Prieto Ramos, F. (2024). Revisiting translator competence in the age of artificial intelligence: The case of legal and institutional translation. *The Interpreter and Translator Trainer, 18*(2), 148–173. https://doi.org/10.1080/1750399X.2024.2344942

Reiter, E. (2018). A structured review of the validity of BLEU. *Computational Linguistics, 44*(3), 393–401. https://doi.org/10.1162/coli_a_00322

Sahari, Y., Qasem, F., Asiri, E., Alasmri, I., Assiri, A., Alqahtani, S., & Mahdi, H. (2024). Evaluating the translation of figurative language: A comparative study of ChatGPT and human translators. *CALR Linguistics Journal*, (15). https://doi.org/10.60149/RTZQ6644

Tiersma, P. M. (1999). *Legal language*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226854809.001.0001

Vilar, D., Freitag, M., Cherry, C., Luo, J., Ratnakar, V., & Foster, G. (2023). Prompting PaLM for translation: Assessing strategies and performance. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 15406–15427). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.859

Wang, Y., Hoang, C., & Federico, M. (2021). Towards modeling the style of translators in neural machine translation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1193–1199). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.94

Yan, J., Yan, P., Chen, Y., Li, J., Zhu, X., & Zhang, Y. (2024). *GPT-4 vs. human translators: A comprehensive evaluation of translation quality across languages, domains, and expertise levels*. arXiv. https://doi.org/10.48550/arXiv.2407.03658

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