Diaa Al-Fekr Journal for Research and Studies

مجلة ضياء الفكر للبحوث والدّراسات

Journal Homepage: https://ojs.diaalfekr.com/index.php/sjlb

Print ISSN: 3006-5356
Online ISSN: 3006-5364

Vol. 1, Issue 9, 2025, pp. 11 – 38

Can Artificial Intelligence Help Bridge the Knowledge Gap?
A Contrastive Study between Translations rendered by AI and Human Translation

هل بإمكان الذكاء الاصطناعي المساعدة في ردم الفجوة المعرفية؟
دراسة مقارنة بين ترجمة الذكاء الاصطناعي وترجمة الانسان

DOI: https://doi.org/10.71090/am13jb30

- Youssef, Muneer Ahmed Ali. (2025). Can Artificial Intelligence Help Bridge the Knowledge Gap? A Contrastive Study between Translations rendered by AI and Human Translation. Diaa Al-Fekr Journal for Research and Studies, Vol. 1, Issue 9, 2025, pp. 11 – 38. https://doi.org/10.71090/am13jb30

Can Artificial Intelligence Help Bridge the Knowledge Gap? A Contrastive Study between Translations rendered by AI and Human Translation

هل بإمكان الذكاء الاصطناعي المساعدة في ردم الفجوة المعرفية؟ دراسة مقارنة بين ترجمة الذكاء الاصطناعي وترجمة الإنسان

د. منير أحمد علي يوسف*

Dr. Muneer Ahmed Ali Youssef *

Abstract:

Recent advancements in artificial intelligence (AI) have catalyzed unprecedented transformations across multiple domains, particularly in translation services, through the advent of sophisticated neural language models (NLMs) and large language models (LLMs). The proliferation of machine translation technologies, including NLM-based platforms such as Google Translate, Microsoft Translator, DeepL, and Amazon Translate, has facilitated cost-efficient and expeditious translation processes. Furthermore, computer-aided translation (CAT) frameworks, including Memoq, Trados, and Memsource, have substantially augmented translation methodologies. The recent emergence of advanced LLMs, notably ChatGPT and GPT-4, has demonstrated remarkable progress in contextual comprehension and textual interaction, approximating human-like engagement with written material.

This study investigates the efficacy of Artificial Intelligence, specifically GPT-4, in translating seminal academic works. The research conducts a comparative analysis of three translations of Chapter Three ("Language and interpretation: philosophical reflections and empirical inquiry") from Noam Chomsky's *New Horizons in the Study of Language and Mind*: (1) an AI-generated translation produced by GPT-4, (2) a translation by three professional translators from Yemen who incorporate GPT-4 into their workflow, and (3) the published Arabic translation by Adnan Hassan, titled "آفاق جديدة في دراسة اللغة والعقل". Through a mixed-methods approach combining automated metrics (BLEU, METEOR, and ROUGE-L) and human evaluation across multiple quality dimensions, this study provides a comprehensive assessment of AI's capabilities and limitations in academic translation.

The findings reveal that while GPT-4 demonstrates impressive capabilities in generating fluent Arabic text and handling general academic language, it struggles with specialized philosophical terminology, cultural adaptation, and preserving nuanced theoretical distinctions. The Human-AI collaborative approach substantially outperforms the GPT-4-only translation across all quality metrics and approaches the quality of the published translation in terms of fluency and readability, while still lagging in terminological consistency and cultural

* جامعة الحديدة/ كلية الآداب – اليمن.

Email: muneer380@gmail.com
* Hodeidah University – Yemen.

adaptation. Notably, the Yemeni translators reported a 40% reduction in translation time when using GPT-4 as a first-draft tool, suggesting significant potential for increasing translation productivity through AI integration.

Keywords: Machine translation, GPT-4, philosophical translation, Arabic translation, human-AI collaboration, Chomsky, academic discourse.

الملخّص:

آذنت التطورات الأخيرة في مجال الذكاء الاصطناعي بتحولات غير مسبوقة في مجالات متعددة، ولا سيما في خدمات الترجمة، وذلك من خلال ظهور نماذج لغوية عصبية (NLM) متطورة والنماذج اللغوية الضخمة (LLM). وقد أدى انتشار تقنيات الترجمة الآلية، التي تشمل المنصات القائمة على النماذج اللغوية العصبية (LLM) مثل مترجم غوغل (Google Translate)، ومترجم مايكروسوفت (Microsoft Translator)، وديب إل (DeepL)، ومترجم أمازون (Amazon Translate)، إلى السماح بإجراء عمليات ترجمة فعالة، قليلة الكلفة وسريعة الإنجاز. كما أسهمت أدوات الترجمة بمساعدة الحاسوب (CAT Tools)، بما في ذلك ميموكيو، وترادوس، وميمسورس، في تعزيز منهجيات الترجمة بشكل ملحوظ. وقد مكنت التطورات الأخيرة في مجال النماذج اللغوية الضخمة المتقدمة، وخاصة تشات جي بي تي (ChatGPT) وجي بي تي-٤ (GPT-4) على سبيل المثال تقدماً ملحوظاً في فهم السياقات المختلفة والقدرة على التفاعل مع النصوص، محاكية في أدائها ومقتربة من القدرات البشرية في التعامل مع النصوص.

يهدف هذا البحث إلى دراسة فعالية الذكاء الاصطناعي وتحديداً جي بي تي-٤ (GPT-4)، وكذلك فعالية دمج أدوات الذكاء الاصطناعي في عمل المترجمين المحترفين في ترجمة الأعمال الأكاديمية الرائدة، من خلال اجراء مقارنة بين ثلاث ترجمات من اللغة الإنجليزية إلى اللغة العربية لأحد فصول كتاب نعوم تشومسكي الموسوم: *آفاق جديدة في دراسة اللغة والعقل* (الفصل الثالث الموسوم اللغة والتفسير: التأملات الفلسفية والاستعلام التجريبي): ١. الترجمة التي قام بها الذكاء الاصطناعي لهذا الفصل المختار، ٢. الترجمة التي قام بها مترجم محترف يستخدم أدوات الذكاء الاصطناعي في انجاز أعماله للفصل المختار، ٣. والترجمة العربية المنشورة لكتاب نعوم تشومسكي "آفاق جديدة في دراسة اللغة والعقل" التي قام بها عدنان حسن للفصل المختار.

تشير نتائج هذه الدراسة أنه على الرغم من أن أنظمة الترجمة المدعومة بالذكاء الاصطناعي قد حققت تقدمًا كبيرًا فيما يتعلق بالدقة والكفاءة، إلا أنها لا تزال تواجه تحديات كبيرة عند التعامل مع تعقيدات الخطاب الأكاديمي. فعلى الرغم من التقدم الذي حققته نماذج مثلGPT-4 ، تظل الخبرة البشرية ضرورية لتمييز الفروق الدقيقة والتفسيرات المعتمدة على السياق. تجادل هذه الدراسة بأن دمج النماذج اللغوية الضخمة المتقدمة مثل GPT-4 وإصداراتها اللاحقة في انجاز عمل المترجمين المحترفين يمكن أن يؤدي إلى تحقيق التآزر لتحقيق الدقة والسرعة وتقليص الكلفة. قد يسهم اتباع مثل هذا النهج التعاوني بين المترجمين المحترفين والذكاء الاصطناعي بشكل كبير في سد الفجوة المعرفية القائمة بين اللغة العربية واللغات الأخرى.

الكلمات المفتاحية: الذكاء الاصطناعي، الترجمة الآلية، الترجمة البشرية، دمج الذكاء الاصطناعي، الفجوة المعرفية، الترجمة الأكاديمية.

Introduction :

In recent decades, the landscape of translation has undergone profound transformations, driven by remarkable advancements in artificial intelligence (AI) technologies. The emergence of sophisticated neural language models (NLMs) and large language models (LLMs) has revolutionized translation processes, offering unprecedented capabilities in cross-linguistic communication. These technological innovations have not only enhanced the efficiency and accessibility of translation services but have also raised important questions about the quality, accuracy, and cultural sensitivity of machine-generated translations, particularly in specialized domains such as academic discourse.

The translation industry has witnessed several evolutionary stages, from rule-based systems to statistical machine translation, and more recently, to neural machine translation (NMT) frameworks. The proliferation of NMT-based platforms such as Google Translate, Microsoft Translator, DeepL, and Amazon Translate has facilitated cost-efficient and expeditious translation processes across multiple languages. Concurrently, computer-aided translation (CAT) frameworks, including Memoq, Trados, and Memsource, have substantially augmented traditional translation methodologies by providing translators with powerful tools to enhance consistency, manage terminology, and improve overall productivity.

The recent emergence of advanced LLMs, notably ChatGPT and GPT-4, represents a significant leap forward in the evolution of translation technologies. These models have demonstrated remarkable progress in contextual comprehension and textual interaction, approximating human-like engagement with written material. Their ability to understand nuanced language, recognize cultural references, and maintain coherence across complex texts suggests potential applications in domains previously considered the exclusive province of human translators, such as literary works, legal documents, and academic publications.

Despite these technological advancements, a significant knowledge gap persists between Arabic and English translations, particularly in academic contexts. The Arabic-speaking world faces a considerable disparity in access to translated literature compared to other language communities. From 1970-1975 to the present, only approximately 10,000 books have been translated into Arabic, a figure equivalent to Spain's annual translation output. This disparity is particularly pronounced in academic and scientific domains, where the timely translation of seminal works is crucial for knowledge dissemination and scholarly advancement.

This study investigates the efficacy of Artificial Intelligence, specifically GPT-4, in translating seminal academic works, with a focus on Noam Chomsky's "New Horizons in the Study of Language and Mind." By conducting a comparative analysis of three distinct translations of Chapter Three ("Language and interpretation: philosophical reflections and empirical inquiry"), this research aims to assess the capabilities and limitations of AI in academic translation, the potential benefits of human-AI collaboration, and how these compare to traditional human translation methods. The translations under analysis include: (1) an AI-generated translation produced by GPT-4, (2) a translation by three professional translators from Yemen who incorporate GPT-4 into their workflow, and (3) the published Arabic translation by Adnan Hassan, titled " *آفاق جديدة في دراسة اللغة والعقل*".

This triangulated approach enables a multifaceted examination of translation quality, accuracy, and cultural appropriateness across different translation methodologies. By employing both quantitative metrics (such as BLEU, METEOR, and ROUGE-L scores) and qualitative assessments (including expert evaluations of terminological precision, stylistic appropriateness, and conceptual accuracy), this study provides a comprehensive evaluation of AI's potential contribution to academic translation.
The significance of this research extends beyond the specific case study to address broader questions about the role of AI in bridging knowledge gaps between languages. By examining how professional translators can leverage advanced LLMs in their workflow, this study explores practical strategies for enhancing productivity, time-efficiency, and terminological consistency in translation processes. The findings may inform future approaches to translation in academic contexts, potentially accelerating the translation of scholarly works and facilitating greater cross-linguistic knowledge transfer.
As AI technologies continue to evolve at a rapid pace, understanding their capabilities and limitations in specialized translation domains becomes increasingly important. This study contributes to this understanding by providing empirical evidence on the current state of AI-assisted translation in academic contexts, while also suggesting future directions for research and practice in this dynamic field.

Literature Review:

1. Evolution of Translation Technologies:

1.1. From Rule-Based to Neural Machine Translation

The field of machine translation has undergone significant evolution over the past several decades. Early rule-based systems of the 1950s and 1960s relied on linguistic rules and dictionaries, producing translations of limited quality and flexibility (Hutchins, 2010). The subsequent development of statistical machine translation (SMT) in the 1990s marked a paradigm shift, utilizing statistical models trained on parallel corpora to generate translations based on probability distributions (Brown et al., 1993). While SMT improved upon rule-based approaches, it still struggled with long-range dependencies and semantic coherence.

The introduction of neural machine translation (NMT) in the mid-2010s represented a revolutionary advancement in the field. Unlike its predecessors, NMT employs artificial neural networks to mimic the human translation process, encoding the source text, decoding it into the target language, and employing attention mechanisms to ensure contextually accurate translations (Bahdanau et al., 2015). This approach has significantly enhanced translation quality by better capturing linguistic nuances and contextual relationships.

1.2. Neural Language Models in Translation

Neural language models (NLMs) have demonstrated remarkable capabilities in improving machine translation through their ability to generalize to long contexts and capture complex linguistic patterns. These models process language as interconnected networks of nodes that learn by passing information between layers and adjusting connections based on training data (Devlin et al., 2019). The transformer architecture, introduced by Vaswani et al. (2017), has become particularly influential in NMT systems due to its parallel processing capabilities and self-attention mechanisms, which allow for more efficient and accurate translations.

Current NMT platforms such as Google Translate, Microsoft Translator, DeepL, and Amazon Translate utilize sophisticated neural networks to provide translations across numerous language pairs. These systems have substantially improved translation accessibility and efficiency, though they continue to face challenges with highly specialized content and culturally nuanced expressions (Läubli et al., 2020).

1.3. Computer-Aided Translation Frameworks

Alongside the development of machine translation systems, computer-aided translation (CAT) frameworks have evolved to enhance human translators' productivity and consistency. Tools such as SDL Trados, MemoQ, and Memsource provide features including translation memory, terminology management, and quality assurance functionalities (O'Brien et al., 2014). These frameworks have become essential in professional translation workflows, allowing translators to leverage previous translations, maintain terminological consistency, and collaborate more effectively on large-scale projects.

CAT tools vary in their capabilities and specializations. SDL Trados is known for its comprehensive features and compatibility with various file formats, while MemoQ offers powerful collaboration features and integration with machine translation engines. Wordfast provides an intuitive interface available as both desktop and online tools, and Phrase (formerly Memsource) employs AI-driven technology to improve translation quality and efficiency (Moorkens et al., 2018). These tools have transformed professional translation practices by automating repetitive tasks and providing translators with resources to enhance their work.

2. Large Language Models and Translation:

2.1. Emergence of Advanced LLMs

The development of large language models (LLMs) such as GPT-3, GPT-4, and similar systems has marked a significant advancement in natural language processing capabilities. These models are trained on vast corpora of text data, enabling them to generate coherent and contextually appropriate content across various domains (Brown et al., 2020). Unlike traditional NMT systems, which are specifically designed and trained for translation tasks, LLMs are general-purpose language models that can perform translation as one of many language-related tasks.

The scale of these models, often comprising billions of parameters, allows them to capture complex linguistic patterns and relationships that smaller models cannot. This scale, combined with sophisticated training techniques such as unsupervised learning and reinforcement learning from human feedback, has enabled LLMs to achieve unprecedented performance on a wide range of language tasks (Ouyang et al., 2022).

2.2. GPT-4's Translation Capabilities

GPT-4, developed by OpenAI, represents one of the most advanced LLMs currently available. Research by Zhu et al. (2023) indicates that GPT-4 has surpassed strong supervised baseline models like NLLB (No Language Left Behind) in 40.91% of translation directions. This performance is particularly noteworthy given that GPT-4 was not specifically optimized for translation tasks, suggesting the emergence of translation as a capability that arises naturally from sufficient scale and training.

GPT-4 demonstrates several unique characteristics in translation contexts. It exhibits resource-efficient acquisition of translation ability, generating moderate translations even for zero-resource languages where parallel training data is scarce or nonexistent. The model also shows interesting patterns in handling instructions, sometimes prioritizing in-context examples over explicit instructions. Additionally, cross-lingual exemplars can provide better task guidance for low-resource translation than exemplars in the same language pairs (Zhu et al., 2023).

Despite these capabilities, GPT-4 still faces challenges in translation, particularly for low-resource languages and specialized domains. It continues to lag behind commercial translation systems like Google Translate in many contexts, highlighting the ongoing need for domain-specific training and refinement (Jiao et al., 2023).

2.3. The Knowledge Gap in Arabic Translation

1. Current State of Arabic Translation

Despite Arabic being the fourth most widely-spoken language worldwide and the official language of 22 Arab states, it faces significant disparities in translation activity compared to other major languages. According to Alshehri et al. (2025), Arabic is ranked only 29th among the top 50 languages for translated literature in the UNESCO Index Translation um database. From 1970-1975 to the present, only approximately 10,000 books have been translated into Arabic, equivalent to Spain's annual translation output. This disparity is particularly evident when comparing translation volumes across languages. The Arab world translates about 330 books annually, merely one-fifth of the number that Greece translates, despite having a significantly larger population (ALTA Language Services, 2020). This limited translation activity contributes to what has been termed a "knowledge gap" between Arab societies and developed countries, a gap that precedes economic or technical disparities (MPRA, 2008).

2. Historical Context and Initiatives

The Arab world has a rich translation history, including the significant Abbasid School of Translation during the Ewan period, which played a pivotal role in preserving and transmitting classical knowledge. Contemporary translation activities in the Arab world can be attributed to four main factors: individual and institutional translations, government-led national translation projects, translation centers, and organizations such as the Arab League for Education, Culture, and Sciences (ALECSO) (Alshehri et al., 2025). Several initiatives have been established to address the translation gap, including the "International Prize of the Custodian of the Two Holy Mosques, King Abdullah ibn Abdul Aziz, for Translation," which aims to honor translators and promote translation from and into Arabic. The United Nations Development Program has also published reports on human development in the Arab world, highlighting that translation creates opportunities for knowledge acquisition and transfer, although it remains a relatively unstructured field (Alshehri et al., 2025).

3. Factors Contributing to the Gap

Multiple factors contribute to the translation gap in Arabic. Institutional censorship and government oversight in the Arab world shape translation choices, often ensuring conformity with cultural norms at the expense of comprehensive knowledge transfer. Limited resources allocated to translation projects and the lack of coordinated translation strategies across the Arab world further exacerbate the issue (Eurozine, 2004). The gap is particularly pronounced in academic and scientific domains, where the timely translation of seminal works is crucial for knowledge dissemination and scholarly advancement. This creates a cycle where limited access to translated academic literature impedes research and educational development, further widening the knowledge gap between Arabic-speaking regions and other parts of the world.

2.4. Potential of AI in Bridging the Knowledge Gap

1. AI-Human Collaboration in Translation

Recent research suggests that collaboration between human translators and AI systems may offer a promising approach to addressing translation challenges. Studies by Läubli et al. (2020) and Daems & Macken (2019) indicate that human-AI collaboration can lead to improvements in translation quality, efficiency, and consistency compared to either human-only or AI-only approaches.

In this collaborative paradigm, AI systems can assist with initial translation drafts, terminology management, and repetitive elements, while human translators provide critical judgment on nuanced meanings, cultural adaptations, and stylistic refinements. This division of labor leverages the respective strengths of both AI and human translators, potentially leading to superior outcomes compared to traditional translation methods.

2. Accelerating Academic Translation

The integration of advanced LLMs like GPT-4 into professional translation workflows holds particular promise for accelerating the translation of academic works. By automating aspects of the translation process, translators may be able to focus their expertise on the most challenging aspects of academic translation, such as specialized terminology, complex theoretical concepts, and discipline-specific conventions (Way, 2018).

This acceleration could potentially increase the volume of translated works available in languages with translation deficits, such as Arabic. By reducing the time and resources required for academic translation, AI-assisted approaches might help address the knowledge gap between Arabic and other languages, facilitating greater cross-linguistic knowledge transfer and scholarly exchange.

3. Challenges and Limitations

Despite its potential, AI-assisted translation faces several challenges in academic contexts. These include the accurate translation of specialized terminology, the preservation of nuanced theoretical arguments, and the adaptation of cultural references and examples (Castilho et al., 2017). Additionally, the rapidly evolving nature of AI technology means that capabilities and limitations are constantly changing, necessitating ongoing evaluation and adaptation of translation practices.

The ethical implications of AI-assisted translation also warrant consideration, including questions about authorship, intellectual property, and the potential homogenization of academic discourse across languages. These challenges highlight the importance of developing thoughtful approaches to AI integration in academic translation that preserve the integrity and diversity of scholarly communication.

3. Methodology:

3.1. Research Design

This study employs a mixed-methods approach to compare three different translations of Chapter Three ("Language and interpretation: philosophical reflections and empirical inquiry") from Noam Chomsky's "New Horizons in the Study of Language and Mind." The research design combines quantitative metrics and qualitative analysis to provide a comprehensive assessment of translation quality across different translation approaches. This triangulated methodology allows for a multifaceted examination of the capabilities and limitations of AI in academic translation, particularly in the context of complex philosophical and linguistic discourse.

3.2. Translation Samples

Three distinct translations of the same source text were analyzed in this study:

1. AI-Generated Translation (GPT-4): A complete translation of Chapter Three produced solely by OpenAI's GPT-4 model without human intervention. The translation was generated by inputting the original English text in segments, with appropriate context provided to maintain coherence across the chapter.

2. Human-AI Collaborative Translation: A translation produced by three professional translators from Yemen who incorporated GPT-4 into their workflow. These translators used GPT-4 as an initial translation tool and then refined, edited, and adapted the output based on their professional expertise and knowledge of the subject matter.

3. Published Human Translation: The published Arabic translation by Adnan Hassan, titled *آفاق جديدة في دراسة اللغة والعقل*, (New Horizons in the Study of Language and Mind). This professionally produced translation, published through established academic channels, serves as a benchmark for comparison.

3.3. Data Collection

The data collection process involved several stages:

1. Source Text Preparation: Chapter Three of Chomsky's "New Horizons in the Study of Language and Mind" was digitized and segmented into

manageable units for translation and analysis. Care was taken to preserve the integrity of complex arguments and ensure that contextual relationships between segments were maintained.

2. AI Translation Generation: The source text segments were input into GPT-4 using the most recent available model at the time of the study. Appropriate prompts were designed to optimize the model's performance for academic translation, including instructions to maintain terminological consistency and preserve the philosophical nuances of the text.

3. Human-AI Collaborative Translation Process: The three Yemeni translators were provided with the source text and access to GPT-4. They documented their workflow, including how they utilized the AI tool, what modifications they made to the AI-generated translations, and what challenges they encountered during the process.

4. Collection of Published Translation: The published Arabic translation by Adnan Hassan was obtained and digitized for comparison. This translation was segmented to align with the same units used for the other two translation types.

3.4. Evaluation Framework

The evaluation framework consisted of both automated metrics and human assessment criteria:

3.4.1 Automated Metrics

Three widely recognized automated metrics were employed to provide quantitative measures of translation quality:

1. BLEU (BiLingual Evaluation Understudy): This metric measures precision by comparing n-grams in the candidate translations with the reference translation (Hassan's published version). BLEU scores range from 0 to 1, with higher scores indicating greater similarity to the reference translation.

2. METEOR (Metric for Evaluation of Translation with Explicit ORdering): Selected for its higher correlation with human judgment, METEOR incorporates stemming and synonymy matching along with exact word matching. This metric is particularly valuable for evaluating translations where meaning preservation is paramount.

3. ROUGE-L (Recall-Oriented Understudy for Gisting Evaluation - Longest Common Subsequence): This metric measures the longest common subsequence between translations, helping to evaluate sentence structure preservation and overall fluency.

3.4.2 Human Assessment Criteria

The human evaluation component employed a structured assessment framework with the following criteria:

1. Accuracy and Meaning Preservation (1-5 scale):
   - Preservation of philosophical concepts and arguments
   - Accuracy of specialized terminology in linguistics and philosophy
   - Maintenance of logical relationships between ideas
   - Preservation of nuanced theoretical distinctions

2. Fluency and Readability (1-5 scale):
   - Grammatical correctness
   - Natural sentence structure and flow
   - Coherence between sentences and paragraphs
   - Stylistic appropriateness for academic discourse

3. Terminological Consistency (1-5 scale):
   - Consistent translation of key terms throughout the text
   - Appropriate use of established Arabic equivalents for philosophical and linguistic terminology
   - Creation of suitable neologisms when necessary
   - Consistency with broader academic conventions in Arabic

4. Cultural and Contextual Adaptation (1-5 scale):
   - Appropriate handling of culturally specific references
   - Adaptation of examples to resonate with Arabic-speaking audiences
   - Sensitivity to cultural norms and expectations
   - Accessibility to the target academic audience

Additionally, evaluators provided qualitative feedback through:

- Identification of specific strengths and weaknesses in each translation
- Documentation of notable translation choices and their implications
- Suggestions for improvement in each translation approach

3.5. Evaluation Process

The evaluation process was conducted in several phases:

1. Preparation Phase:

- Selection of 50 representative passages from Chapter Three, ensuring coverage of various philosophical arguments, linguistic concepts, and stylistic features
- Development of evaluation templates and scoring guidelines
- Training of evaluators to ensure consistent application of assessment criteria

2. Blind Evaluation:

- Evaluators were presented with the three translations in randomized order, without information about which translation method was used
- Each evaluator assessed all 50 passages across all three translations
- Evaluations were conducted independently to prevent bias

3. Automated Metric Calculation:

- BLEU, METEOR, and ROUGE-L scores were calculated for each translation
- The published translation by Adnan Hassan was used as the reference for these calculations
- Scores were computed both for the entire chapter and for each of the 50 selected passages

4. Comparative Analysis Workshop:

- After individual evaluations were complete, evaluators participated in a workshop to discuss their findings
- Specific examples illustrating key strengths and weaknesses were identified
- Patterns across evaluations were noted and analyzed

3.6. Data Analysis

The data analysis process involved both quantitative and qualitative components:

1. Statistical Analysis:

- Calculation of mean scores and standard deviations for each human assessment criterion across the three translation types
- Comparative analysis of automated metric scores
- Correlation analysis between human assessments and automated metrics
- Significance testing to determine meaningful differences between translation types

2. Thematic Analysis:

- Coding of qualitative feedback to identify recurring themes

- Categorization of strengths and weaknesses for each translation type
- Identification of patterns in translation approaches and outcomes
- Development of a conceptual framework for understanding AI's role in academic translation

3. Error Analysis:
- Classification of translation errors by type (e.g., terminological, syntactic, semantic)
- Comparison of error patterns across the three translation types
- Identification of systematic issues in AI-generated translations
- Analysis of how human intervention addressed or failed to address AI translation errors

3.7. Ethical Considerations

Several ethical considerations were addressed in the design and implementation of this study:

1. Informed Consent: All human translators and evaluators provided informed consent for their participation in the study. They were fully briefed on the research objectives, methodologies, and how their contributions would be used.

2. Acknowledgment of Biases: Potential biases in evaluation criteria were acknowledged and addressed through the use of multiple evaluators, blind assessment procedures, and a combination of quantitative and qualitative measures.

3. Transparency: The limitations of the methodology, including the focus on a single chapter and the rapidly evolving nature of AI technology, were explicitly acknowledged in the reporting of results.

4. Fair Representation: Care was taken to represent all three translation approaches fairly, without privileging any particular method or making unwarranted claims about their relative merits.

3.8. Limitations

This study acknowledges several limitations that should be considered when interpreting the results:

1. Scope: The focus on a single chapter from one academic work limits the generalizability of findings to other types of academic texts or subject matters.

2. Technological Currency: The rapidly evolving nature of AI technology means that the capabilities of GPT-4 and similar models will continue to develop, potentially rendering some findings outdated as new models emerge.

3. Evaluator Subjectivity: Despite efforts to ensure objectivity, human evaluations inevitably contain an element of subjectivity that may influence the assessment of translation quality.

4. Language Pair Specificity: The findings may be specific to the English-Arabic language pair and may not generalize to other language combinations without further research.

5. Cultural Context: The cultural and academic context of Arabic translation may influence the reception and evaluation of translations in ways that differ from other linguistic and cultural contexts.

Despite these limitations, this study provides valuable insights into the current capabilities and limitations of AI in academic translation, as well as the potential benefits of human-AI collaboration in addressing the knowledge gap between Arabic and English academic literature.

4. Results

4.1. Automated Metrics Analysis

4.1.1 Overall Comparison

The quantitative analysis of the three translation types using automated metrics revealed notable patterns in translation quality. Table 1 presents the overall scores for BLEU, METEOR, and ROUGE-L across the three translation types for the complete Chapter Three.

Table 1: Automated Metric Scores for Complete Chapter

Translation Type	BLEU	METEOR	ROUGE-L
GPT-4 Only	0.42	0.61	0.58
Human-AI Collaborative	0.68	0.79	0.74
Published (Hassan)	1.00	1.00	1.00

The Human-AI collaborative translation achieved substantially higher scores than the GPT-4-only translation across all three metrics, with particularly significant improvements in BLEU scores. This suggests that human intervention substantially enhanced the lexical and structural alignment with the reference translation. ROUGE-L scores, however, showed the smallest gap (0.16) between the two AI-involved translations. METEOR scores, which incorporate synonymy and stemming, also demonstrated a comparatively small difference (0.18), indicating that GPT-4 demonstrated reasonable semantic accuracy even without human refinement.

4.1.2 Passage-Level Analysis

When examining the 50 selected representative passages individually, we observed variation in performance across different types of content.

Passages containing specialized philosophical terminology (passages 3, 7, and 12) showed the largest disparity between GPT-4-only and Human-AI collaborative translations. In these instances, human expertise in domain-specific terminology appeared to provide significant value. Conversely, passages with more general linguistic descriptions (passages 2, 5, and 9) showed smaller gaps between the two AI-involved approaches, suggesting that GPT-4 handled general academic language more effectively than specialized philosophical discourse.

4.1.3 Correlation Between Metrics

Correlation analysis between the three automated metrics revealed strong positive correlations (Pearson's r > 0.85 for all pairs), indicating general agreement in their assessment of translation quality. However, METEOR showed the strongest correlation with human evaluations (r = 0.79), followed by ROUGE-L (r = 0.72) and BLEU (r = 0.65). This aligns with previous research suggesting that METEOR's incorporation of semantic features makes it more reflective of human judgment in translation quality assessment.

4.2. Human Evaluation Results

4.2.1 Overall Quality Assessment

The human evaluation provided a more nuanced understanding of translation quality across the three approaches. Table 2 presents the mean scores (on a 1-5 scale) for each assessment criterion.

Table 2: Mean Human Evaluation Scores

Translation Type	Accuracy & meaning	Fluency & Readability	Terminological Consistency	Cultural Adaptation
GPT-4	3.2	3.8	2.9	2.5

Only				
Human-AI Collaborative	4.3	4.5	4.2	3.9
Published (Hassan)	4.7	4.6	4.8	4.7

The Human-AI collaborative translation approached the quality of the published translation in terms of fluency and readability, with a mean score only 0.1 points lower. This suggests that the combination of AI-generated text and human refinement produced highly readable academic prose in Arabic. However, the collaborative approach still lagged behind the published translation in accuracy, terminological consistency, and cultural adaptation, indicating areas where professional translation expertise provides value beyond what can be achieved through post-editing of AI output.

The GPT-4-only translation performed best in fluency and readability (3.8), demonstrating the model's strong capabilities in generating grammatically correct and coherent Arabic text. However, it scored notably lower in terminological consistency (2.9) and cultural adaptation (2.5), highlighting specific weaknesses in AI-only approaches to academic translation.

4.2.2 Evaluator Agreement

Inter-rater reliability analysis showed strong agreement among evaluators for accuracy and terminological consistency (Krippendorff's Î± > 0.80), moderate agreement for fluency and readability (Î± = 0.72), and lower agreement for cultural adaptation (Î± = 0.65). The lower agreement on cultural adaptation reflects the subjective nature of assessing cultural appropriateness and suggests that this dimension of translation quality may be particularly challenging to evaluate consistently.

4.2.3 Qualitative Feedback

Thematic analysis of evaluators' qualitative feedback revealed several recurring patterns across the three translation types:

GPT-4-Only Translation

Strengths:

- Impressive handling of complex sentence structures
- Generally accurate translation of core concepts
- Consistent internal style throughout the chapter
- Good preservation of logical flow between arguments

Weaknesses:

- Inconsistent translation of key philosophical terms
- Literal translation of culturally specific examples
- Occasional misinterpretation of nuanced theoretical distinctions
- Limited adaptation to Arabic academic conventions

One evaluator noted: "The AI translation demonstrates remarkable fluency and captures the general thrust of Chomsky's arguments, but it struggles with the philosophical nuances that require deep domain knowledge. Terms like 'intentionality' and 'referential opacity' are translated inconsistently, sometimes changing meaning within the same paragraph."

Human-AI Collaborative Translation

Strengths:

- Significant improvement in terminological consistency over AI-only translation
- Better preservation of philosophical nuances
- More natural integration of examples for Arabic readers
- Appropriate use of established Arabic philosophical vocabulary

Weaknesses:

- Occasional stylistic inconsistencies between sections
- Some instances of excessive literalness in complex arguments
- Less elegant phrasing than the published translation
- Variable quality in cultural adaptation

An evaluator commented: "The collaborative translation successfully addresses many of the terminological issues present in the AI-only version. The human translators clearly recognized when GPT-4 misunderstood philosophical concepts and made appropriate corrections. However, the result sometimes lacks the stylistic elegance and cultural resonance of Hassan's translation."

4.3. Error Analysis

4.3.1 Error Types and Distribution

Detailed error analysis revealed distinct patterns across the three translation types. Figure 2 illustrates the distribution of error types as a percentage of total errors identified.

The GPT-4-only translation showed the highest proportion of terminological errors (42% of total errors), followed by meaning distortion errors (27%), cultural appropriateness errors (18%), and grammatical/stylistic errors (13%). This distribution highlights the model's primary challenges in handling specialized academic vocabulary and preserving precise meanings in philosophical discourse.

The Human-AI collaborative translation showed a more balanced error distribution, with terminological errors reduced to 24% of total errors. Meaning distortion errors (31%) became the most common category, suggesting that while human intervention successfully addressed many terminological issues, some subtle meaning distortions remained challenging to detect and correct. Cultural appropriateness errors (26%) and grammatical/stylistic errors (19%) made up the remainder.

The published translation had the fewest overall errors, with meaning distortion being the most common type (45% of a much smaller total), followed by cultural appropriateness (30%), terminological errors (15%), and grammatical/stylistic errors (10%).

4.3.2 Specific Error Examples

Analysis of specific error instances provided insights into the nature of translation challenges across the three approaches:

1. Terminological Inconsistency (GPT-4-Only):

In a passage discussing "rule-following," GPT-4 translated the term using three different Arabic expressions within the same section, creating confusion about whether Chomsky was referring to the same or different concepts.

2. Meaning Distortion (Human-AI Collaborative):

In explaining Chomsky's critique of Quine's indeterminacy thesis, the collaborative translation reversed the logical relationship between two key points, significantly altering the argument's structure despite using correct terminology.

3. Cultural Adaptation (GPT-4-Only):

When translating an example involving American cultural references, GPT-4 produced a literal translation that failed to resonate with Arabic readers, whereas both the collaborative and published translations substituted culturally appropriate analogies.

4. Stylistic Appropriateness (All Translations):

All three translations occasionally struggled with Chomsky's complex sentence structures, but employed different strategies: GPT-4 often preserved the original structure even when awkward in Arabic; the collaborative translation sometimes broke sentences into shorter units; and Hassan's translation more confidently restructured sentences to align with Arabic stylistic preferences.

4.4. Comparative Strengths Analysis

4.4.1 Domain-Specific Performance

Analysis of performance across different content domains revealed interesting patterns in the relative strengths of each translation approach:

1. General Linguistic Descriptions:
GPT-4 performed relatively well (mean accuracy score: 3.7) in translating general descriptions of linguistic phenomena, approaching the quality of the collaborative translation (4.2) and the published translation (4.6) in these sections.

2. Philosophical Argumentation:
All translations showed lower scores in highly abstract philosophical arguments, but with significant differences: GPT-4 (2.8), collaborative (3.9), published (4.5). This domain showed a significant gap between AI-only and human-involved translations.

3. Technical Terminology:
The collaborative translation showed its greatest improvement over GPT-4 in handling technical terminology, with accuracy scores of 4.3 versus 2.7, approaching the published translation's 4.7. This domain also exhibited the largest disparity (1.6 points) between the GPT-4-only and Human-AI collaborative translations.

4.4.2 Workflow Efficiency

Analysis of the Human-AI collaborative translation process revealed significant efficiency gains compared to traditional translation methods. The Yemeni translators reported that using GPT-4 as a first-draft tool reduced their overall translation time by approximately 40% compared to their usual workflow. However, they noted that the post-editing process for philosophical content required substantial effort, with an average of 65% of the AI-generated text requiring some form of revision.

The translators identified several patterns in their editing process:
- Terminology correction was the most time-consuming aspect
- Restructuring sentences for Arabic stylistic preferences was frequently necessary
- Adding explanatory phrases for culturally specific concepts was often required
- Standardizing terminology across the chapter required careful attention

One translator noted: "GPT-4 provided an excellent starting point that saved considerable time in producing the initial draft. However, the post-editing process required deep subject matter knowledge and careful attention to terminological consistency. The AI excelled at generating grammatically correct Arabic but needed significant human intervention to achieve the precision required for philosophical discourse."

4.5. Summary of Key Findings

The results reveal several key insights about the relative strengths and limitations of AI-only, Human-AI collaborative, and traditional human translation approaches for academic content:

1. The Human-AI collaborative approach substantially outperformed the GPT-4-only translation across all quality metrics, demonstrating the value of human expertise in refining AI-generated translations.

2. The collaborative approach achieved comparable results to the published translation in terms of fluency and readability but still lagged in accuracy, terminological consistency, and cultural adaptation.

3. GPT-4 demonstrated impressive capabilities in generating fluent Arabic text and handling general academic language but struggled with specialized philosophical terminology, cultural adaptation, and nuanced theoretical distinctions.

4. The efficiency gains reported by translators in the collaborative approach suggest potential for increasing translation productivity, particularly for academic content where specialized knowledge is required.

5. Different content types showed varying levels of success across translation approaches, with general linguistic descriptions being most amenable to AI translation and abstract philosophical arguments presenting the greatest challenges.

These findings provide empirical evidence for both the current capabilities and limitations of AI in academic translation, as well as the potential benefits of integrating AI tools into professional translators' workflows.

5. Discussion:

5.1 Implications of AI-Assisted Translation for Academic Discourse

The comparative analysis of the three translation approaches, GPT-4 only, Human-AI collaborative, and traditional human translation reveals significant implications for the future of academic translation, particularly in bridging knowledge gaps between languages. The findings demonstrate that while AI has made remarkable progress in translation capabilities, the integration of human expertise with AI tools currently offers the most promising approach for translating complex academic works.

The performance of GPT-4 in generating fluent Arabic text without human intervention represents a significant advancement in machine translation technology. The model's ability to produce grammatically correct sentences and maintain logical flow throughout a complex philosophical text demonstrates how far AI translation has evolved from earlier rule-based and statistical approaches. This level of fluency would have been unimaginable just a few years ago and suggests that AI systems are increasingly capable of handling sophisticated linguistic tasks. However, the consistent gaps in terminological precision, cultural adaptation, and philosophical nuance highlight the continuing limitations of even the most advanced AI systems when dealing with specialized academic discourse.

The Human-AI collaborative approach emerged as a particularly promising model for academic translation. By combining GPT-4's capabilities in generating initial drafts with human expertise in terminology, cultural context, and subject matter knowledge, this approach achieved results that significantly outperformed AI-only translation and approached the quality of professional human translation in several dimensions. The reported 40% reduction in translation time suggests substantial efficiency gains, which could potentially increase the volume of academic works translated into languages with translation deficits, such as Arabic. This finding aligns with research by Läubli et al. (2020) and Daems & Macken (2019), who found that human-AI collaboration can lead to improvements in translation quality, efficiency, and consistency compared to either human-only or AI-only approaches.

The superior performance of Adnan Hassan's published translation across most quality metrics, particularly in terminological consistency and cultural adaptation, reinforces the continuing value of specialized human expertise in academic translation. Hassan's deep understanding of both Chomsky's theoretical framework and Arabic philosophical discourse enabled translations that not only preserved meaning but also resonated with the target audience in ways that AI-assisted approaches could not fully match. This suggests that while AI tools may augment human translation processes, they are unlikely to replace the need for subject matter expertise and cultural knowledge in the foreseeable future, especially for works of philosophical complexity.

5.2 Bridging the Knowledge Gap: Potential and Limitations

The findings of this study have significant implications for addressing the knowledge gap between Arabic and English academic literature. As documented by Alshehri et al. (2025), the limited number of translations available in Arabic approximately 330 books annually compared to Greece's five times larger output represents a substantial barrier to knowledge transfer and academic development in the Arab world. The efficiency gains demonstrated by the Human-AI collaborative approach suggest a potential pathway for increasing translation volume without compromising quality beyond acceptable limits.

For academic institutions, publishers, and translation centers in the Arab world, the integration of AI tools like GPT-4 into professional translation workflows could potentially accelerate the translation of seminal works across various disciplines. The 40% reduction in translation time reported by the Yemeni translators in this study, if generalizable, could significantly increase translation output with the same human resources. However, the substantial post-editing requirements, particularly for specialized terminology and cultural adaptation, indicate that this approach still demands significant human expertise and cannot be viewed as a fully automated solution.

The varying performance across different content types also suggests that a strategic approach to AI integration might be most effective. For sections of academic texts dealing with general descriptions or standard academic language, AI tools might require minimal human intervention, allowing translators to focus their expertise on more challenging sections involving specialized terminology, abstract philosophical concepts, or culturally nuanced arguments. This targeted application of human expertise could optimize the efficiency gains while maintaining acceptable quality standards.

However, several limitations must be acknowledged in considering AI's potential to bridge the knowledge gap. First, the focus on a single chapter from one academic work limits the generalizability of findings to other academic disciplines. Scientific, legal, or historical texts might present different challenges for AI translation. Second, the rapidly evolving nature of AI technology means that capabilities are constantly changing, requiring ongoing evaluation and adaptation of translation practices. Third, the infrastructure and training required to implement effective Human-AI collaborative workflows may present barriers in regions with limited technological resources or specialized expertise.

5.3 Quality Dimensions in Academic Translation

The multidimensional evaluation framework employed in this study revealed important insights about the relative importance of different quality dimensions in academic translation. While automated metrics provided valuable quantitative comparisons, the human evaluation highlighted nuances that numerical scores alone could not capture.

The finding that METEOR showed a strong correlation with human evaluations (r = 0.79) supports previous research suggesting that metrics incorporating semantic features better reflect human judgment in translation quality assessment. This has methodological implications for future studies of academic translation quality, suggesting that METEOR might be preferred over BLEU for evaluating translations of complex academic texts where meaning preservation is paramount. The human evaluation revealed that different quality dimensions presented varying levels of challenge for AI-assisted translation. The relatively strong performance in fluency and readability across all translation types suggests that modern AI systems have largely overcome the grammatical and structural challenges that plagued earlier machine translation approaches. However, the persistent gaps in terminological consistency, cultural adaptation, and preservation of philosophical nuance highlight areas where human expertise remains essential.

The error analysis provided further insights into the specific challenges of translating philosophical discourse. The high proportion of terminological errors in the GPT-4-only translation (42% of total errors) underscores the importance of domain-specific knowledge in academic translation. Terms like "rule-following," "intentionality," and "referential opacity" carry specific meanings within philosophical discourse that require precise and consistent translation to preserve the integrity of the arguments. The reduction of terminological errors to 24% in the Human-AI collaborative translation demonstrates how human expertise can effectively address this limitation of AI systems.

The challenges in cultural adaptation, particularly evident in the translation of examples and analogies, highlight the importance of cultural knowledge in effective academic translation. While GPT-4 demonstrated some awareness of cultural differences, its literal translations of culturally specific references often failed to resonate with Arabic readers. Both the collaborative and published translations showed greater success in substituting culturally appropriate analogies, suggesting that cultural adaptation remains an area where human judgment is particularly valuable.

5.5 Ethical Considerations and Future Directions

The integration of AI into academic translation raises several ethical considerations that merit further exploration. Questions of authorship and attribution become increasingly complex when translations result from human-AI collaboration. While the human translators in this study maintained clear editorial control over the final product, the substantial contribution of GPT-4 to the initial draft raises questions about how such contributions should be acknowledged in academic contexts.

Looking to the future, several promising research directions emerge from this study:

1. Longitudinal Studies: As AI technology continues to evolve rapidly, longitudinal studies tracking improvements in academic translation capabilities over time would provide valuable insights into the changing relationship between human and AI contributions.
2. Discipline-Specific Investigations: Expanding similar comparative analyses to other academic disciplines, such as natural sciences, social sciences, or legal studies, would help identify whether the patterns observed in philosophical translation generalize to other domains.
3. Translator Experience: Research into how translators of varying experience levels interact with AI tools could provide insights into optimal training and workflow design for Human-AI collaborative translation.
4. Cultural Adaptation: Further investigation into how AI systems can be better trained to handle cultural adaptation in academic translation would address one of the key limitations identified in this study.
5. Reader Reception: Studies examining how readers perceive and comprehend AI-assisted translations compared to traditional human translations would provide important insights into the real-world impact of these different approaches.

In conclusion, the findings of this study suggest that while AI has made remarkable progress in translation capabilities, the most promising approach for academic translation currently lies in thoughtful integration of AI tools with human expertise. By leveraging the respective strengths of AI systems and human translators, the Human-AI collaborative approach offers potential for both increasing translation volume and maintaining acceptable quality standards, potentially contributing to narrowing the knowledge gap between languages like Arabic and English. However, realizing this potential will require ongoing research, thoughtful implementation strategies, and continued recognition of the irreplaceable value of human expertise in specialized domains of translation.

6. Conclusion:

This study has examined the efficacy of Artificial Intelligence, specifically GPT-4, in translating complex academic works by conducting a comparative analysis of three translations of Chapter Three from Noam Chomsky's "New

Horizons in the Study of Language and Mind." Through a mixed-methods approach combining automated metrics and human evaluation, we have gained valuable insights into the capabilities and limitations of AI in academic translation, as well as the potential benefits of integrating AI into professional translators' workflows.

The findings reveal a nuanced picture of AI's current role in academic translation. While GPT-4 demonstrated impressive capabilities in generating fluent Arabic text and handling general academic language, it struggled with specialized philosophical terminology, cultural adaptation, and preserving nuanced theoretical distinctions. These limitations highlight the continuing importance of human expertise in academic translation, particularly for works involving complex philosophical concepts and culturally specific references.

The Human-AI collaborative approach emerged as a promising middle ground, substantially outperforming the GPT-4-only translation across all quality metrics and approaching the quality of the published professional translation in terms of fluency and readability. The reported 40% reduction in translation time suggests significant potential for increasing translation productivity through AI integration. However, the substantial post-editing requirements, particularly for specialized terminology and cultural adaptation, indicate that this approach still demands significant human expertise and cannot be viewed as a fully automated solution.

References:

1. ALTA Language Services. (2020). A Note on Arabic Literacy and Translation. Beyond Words.

2. Alshehri, F. A., Alshehri, F. A., Qassem, M., Bedjaoui, W., & Alfaisal, A. A. (2025). The dynamics of translation from and into Arabic in the Arab world: bibliometric analysis of the Index Translationum UNESCO database (1979-2012). Humanities and Social Sciences Communications, 43.

3. Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. 3rd International Conference on Learning Representations.

4. Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization.

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

6. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

7. Bulatova, G. (2024). Metrics for evaluation of translation accuracy. Trusted Data Science @ Haleon.

8. Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? The Prague Bulletin of Mathematical Linguistics, 108(1), 109-120.

9. Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.

10. Chomsky, Noam (2009). *New Horizons in the Study of Language and Mind*. translated by: Adnan Hassan, 1st Edition. Lattakia – Syria Dar Al-Hiwar for Publishing and Distribution.

11. Daems, J., & Macken, L. (2019). *Interactive adaptive MT versus post-editing: a comparative study*. Translation Technology Research in Interaction with Translation Studies.

12. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019.

13. Eurozine. (2004). Arab books and human development.

14. Hutchins, J. (2010). Machine translation: A concise history. Journal of Translation Studies, 13(1-2), 29-70.

15. Jiao, W., Wang, W., Xie, J., Feng, S., & Liu, Y. (2023). Is ChatGPT a good translator? A preliminary study. arXiv preprint arXiv:2301.08745.

16. Läubli, S., Castilho, S., Neubig, G., Sennrich, R., Shen, Q., & Toral, A. (2020). A set of recommendations for assessing human-machine parity in language translation. Journal of Artificial Intelligence Research, 67, 653-672.

17. Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. Text Summarization Branches Out.

18. Lommel, A. R., Burchardt, A., & Uszkoreit, H. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. TradumÃ tica: tecnologies de la traducciÃ³, (12), 455-463.

19. Moorkens, J., Castilho, S., Gaspari, F., & Doherty, S. (Eds.). (2018). Translation quality assessment: From principles to practice. Springer.

20. MPRA. (2008). Performance of the Arabic Book Translation Industry in Selected Arab Countries.

21. O'Brien, S., Simard, M., & Specia, L. (2014). Machine translation and the challenge of human language technology. Machine Translation, 28(3-4), 141-143.

22. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

23. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.

24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

25. Way, A. (2018). Quality expectations of machine translation. In Translation Quality Assessment (pp. 159-178). Springer.

26. Zhu, W., Liu, H., Dong, Q., Xu, J., Huang, S., Kong, L., ... & Li, L. (2023). Multilingual machine translation with large language models: Empirical results and analysis. arXiv preprint arXiv:2304.04675.

Funding: This research received grant no. (636/2024) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.