Journal of Translation and Language Studies
https://jtls.sabapub.com

2025 Volume 6, Issue 2: 40–54

DOI: https://doi.org/10.48185/jtls.v6i3.1945

AI-Driven and Large Language Models-Based Translation of Arabic News Texts into English: A Comparative Evaluation

Abdalwahid Noman [1*], Najeeb Almansoob [2], Othman Saleh Mohammed[3], Yasser Alrefaee [4]

[1&2]Al Janad university for Science and Technology, Yemen
[3]University of Saba Region, Yemen.
[4]Albaydha University, Yemen

Received: 09.10.2025 • Accepted: 04.12.2025 • Published: 20.12.2025 • Final Version: 31.12.2025

Abstract: The proliferation of artificial intelligence (AI) has profoundly reshaped machine translation, particularly through the advent of Large Language Models (LLMs). This study provides a systematic comparative evaluation of three prominent AI-driven translation tools (Google Translate, Reverso, Yandex) and three state-of-the-art LLMs (ChatGPT-4, Gemini-1.5-Pro, Bing) for translating Arabic news texts into English. Employing a quantitative research design, a corpus of twenty diverse Arabic news articles from major outlets was compiled. Expert-validated human translations served as benchmarks. Translation outputs were analyzed using a three-tiered framework: (1) classification of errors into lexico-semantic, syntactic, and formatting types; (2) performance assessment via a five-point scoring rubric; and (3) determination of accuracy levels. Results reveal that lexico-semantic errors were the most prevalent (45.22%), followed by formatting (32.27%) and syntactic errors (22.50%). Among all systems, ChatGPT-4 demonstrated superior performance, committing the fewest total errors (19 out of 471) and achieving the highest mean accuracy score (7.68/8.00), with 75% of its outputs rated as "highly accurate." In stark contrast, the AI-driven tool Reverso performed least effectively, recording the highest error count (128) and the lowest mean score (5.94/8.00). The findings establish a clear performance hierarchy, indicating that LLMs, especially ChatGPT-4, significantly outperform traditional AI-driven tools in handling the linguistic and contextual complexities of Arabic news translation. However, persistent error patterns underscore the continued necessity for human post-editing to ensure precision in professional and media-specific translation contexts.

Key Words: Machine Translation, Large Language Models (LLMs), Arabic-English Translation, News Media, Error Analysis, Translation Accuracy, ChatGPT-4, Comparative Evaluation

1. Introduction

The rapid advancement of artificial intelligence (AI) has triggered significant transformations across various fields of knowledge, and the translation field is no exception. Just as the internet revolutionized communication and information exchange in the late 20th century, AI translation tools are transforming the ways in which languages are processed, interpreted, and understood. According to Falempin and Ranadireksa (2024), these tools have introduced new methods that enhance global communication and mutual understanding. Similarly, Siu (2024) noted that these AI-driven techniques have led to the development of new techniques that are capable of translating languages instantly and supporting real-time communication.

AI-driven translation tools and large language models (LLMs) have introduced significant improvements in translation quality, fluency, and contextual accuracy (Chen, et al. 2024). Deng (2016) pointed out that these systems utilize neural networks and deep learning techniques to process vast amounts of linguistic data, enabling more refined and accurate outputs over time. Farghal and Haider

(2024) emphasized that the emergence of LLMs such as GPT-3, Chat GPT, and Gemini-pro has significantly enhanced the capabilities of machine translation by incorporating deeper contextual understanding. In the same context, Mohsen (2024) argued that these models outperform traditional machine translation systems in capturing contextual meaning. Ravshanovna (2024, P. 639) further supported this view, stating that AI-driven translation systems "provide translations that are contextually richer and more accurate than those produced by traditional CAT tools". However, despite these advancements, concerns persist regarding their reliability and accuracy, particularly when applied to complex or domain-specific content such as legal and literary texts (Zanaty 2024).

In contrast, traditional online translation services— while useful as aids—have well-documented limitations. Previous studies indicate that these systems often fail to adequately convey linguistic and cultural nuances, domain-specific terminology, and idiomatic expressions. Human intervention is frequently required to ensure clarity, coherence, and readability. For instance, Sholikhah et al. (2021) emphasized that machine translation systems struggle with culturally embedded language. Chacha and Mwangi (2024) further found that while these tools perform well in basic semantic translations, they lack accuracy when translating idiomatic or culturally nuanced expressions. Similarly, Abdelaal and Alazzawie (2020) observed that machine translation systems frequently distort idiomatic expressions, leading to misleading or incorrect translations. Additionally, Zinhom (2024) noted that despite their advantages, machine translation tools face considerable challenges in translating colloquial Arabic, particularly as it appears in contemporary mass media and literature.

In today's globalized media landscape, the need for accurate and efficient translation has intensified, as language precision is essential for effective communication. As a result, machine translation has become increasingly significant, as news agencies and journalists often rely on these tools for swift and cost-effective multilingual content dissemination. However, the translation of media texts, in particular, presents a distinct set of challenges. Ahmed (2024) affirmed that precision in media translation is important to preserve the integrity of news reporting and avoid misunderstandings. Shafia (2021) argues that news translation is more complex than conventional translation due to connotation, perspective, and ideology, with editing and adaptation often altering meaning across languages and cultures.

Although significant advancements have been made in AI-driven translation tools, these systems still face several notable challenges, particularly when translating complex and culturally sensitive Arabic media texts. Given these challenges, the problem of the study centers on the translation accuracy, as these tools frequently exhibit errors that can significantly affect the quality of the content. The researchers have observed frequent errors in translated media texts on websites, particularly by journalists or social media bloggers. These errors mainly reflect their reliance on machine translation; the issue motivated the researchers to investigate the performance of six AI-driven translation tools— Google Translate, Reverso, and Yandex, and large language models— Chat GPT-4, Bing, and Gemini-1.5-Pro. Despite the importance of this problem, no research has extensively examined a variety of tools, assessed their performance and translation accuracy in translating Arabic news texts into English. Therefore, there is a pressing need to bridge this gap and find out which of these tools performs best in translating Arabic news texts into English. The study thus seeks to answer the following questions:

1- What are the most common types of errors made by AI-driven translation tools such as Google Translate, Reverso, Yandex, and large language models like Chat GPT-4, Bing, and Gemini-1.5-Pro when translating Arabic news texts into English?
2- Which of the investigated translation tools exhibits the lowest frequency of errors and demonstrates better performance in translating Arabic news texts into English?
3- To what extent does translation accuracy vary among the studied tools when translating Arabic news texts into English?

Furthermore, the significance of this study springs from its focus on the application of AI-based translation tools to the media domain translation. Media texts pose unique translation challenges due to their concise style, cultural references, idiomatic expressions, and evolving terminology, all of which cannot be adequately captured by the examined translation tools. The findings from this research provide significant insights for multiple stakeholders— including translators, journalists, teachers, students, and researchers— by helping them select the most effective tools based on performance. Consequently, this study aims to achieve the following objectives:

1- To identify and categorize the common types of errors made by AI-driven translation tools—Google Translate, Reverso, Yandex, and large language models—Chat GPT-4, Bing, and Gemini-1.5-Pro, when translating Arabic news texts into English.
2- To evaluate the performance of the examined translation tools in terms of error rates and translation accuracy and identify which tools perform better in translating Arabic news texts into English.
3- To compare between the translation outputs made by the examined tools in terms of accuracy when translating Arabic news into English.

2. Literature Review

The evaluation of the performance of AI-driven translation tools and LLMs and their translation accuracy is a multifaceted topic that involves linguistic analysis, and contextual understanding, grounded in the implementation of clearly structured evaluation criteria. The complexity of Arabic's morphology and syntax presents unique challenges for translation systems, necessitating advanced models and evaluation techniques to ensure accuracy. Various studies have studied different automated translation tools and evaluated the quality of their translation outputs. This literature review synthesizes findings from several studies to provide insights into the current state of AI-driven and large language models translation technologies and their performance.

Chandra et al. (2025) evaluated the performance of large language models (LLMs) and Google Translate in translating selected Indian languages into English. The study focused on the sentiment and semantic accuracy of translations produced by GPT-3.5, GPT-4o, and Gemini in comparison with expert human translations. The researchers adopted a comparative analysis approach, incorporating both sentiment and semantic assessment techniques to measure translation quality. Their findings revealed that while LLMs have made significant advancements in handling low-resource languages, challenges persist in preserving sentiment and semantic nuances, GPT-4o and GPT-3.5 showed a higher degree of sentiment preservation than Google Translate, especially in the translation of philosophical and religious texts.

Tekgurler (2025) explored the translation capacity of Gemini in processing historical, low-resourced language texts, focusing on an 18th-century Ottoman Turkish manuscript. Using qualitative textual analysis, the study investigated how safety mechanisms embedded within AI models affect translation accuracy. The research revealed that 14–23% of the manuscript was flagged as harmful content, leading to partial or failed translation outputs and the current LLMs face limitations in translating contextually dense, emotionally charged content due to algorithmic restrictions and ethical filtering.

Sidiya et al. (2024) conducted a comprehensive review of Arabic-English machine translation models, analyzing CNNs, LSTMs, NMT, BERT, and hybrid Transformer-CNN architectures. The study built and tested LSTM and BERT-based models, providing a comparative analysis of their translation performance. The findings of the research emphasized the need for high-quality datasets and standardized evaluation benchmarks to improve Arabic-English translation accuracy.

Almaaytah and Almahasees (2024) investigated the quality of artificial intelligence translation for special needs terms from English into Arabic. The study data were taken from five movies made for people with special needs. The study analyzed data using the error analysis framework of Costa et al. To highlight the strengths and weaknesses of the two tools, the study found that two systems made frequent errors in lexis: semantics, grammar, and orthography.

Al-Salman and Haider (2024) conducted an empirical evaluation of Google Translate, Gemini-pro, and Chat GPT in translating Arabic research titles from the humanities and social sciences into English. Using Koponen’s (2010) translation error strategy framework, the study found that translations were commonly marked by syntactic and sense-related errors, particularly in rendering polysemous terms. Among the three tools, Gemini-pro demonstrated the highest translation accuracy, whereas Google

Translate and Chat GPT exhibited the most equivalence-based errors. Interestingly, human translations contained the fewest diction errors but had the highest number of syntactic inaccuracies, suggesting challenges in target language proficiency.

Jiang et al., (2023) investigated the distinguishability of human translations, neural machine translation (NMT), and Chat GPT-generated translations through linguistic and statistical analysis. Employing machine learning classifiers and multidimensional analysis (MDA), the study found that Chat GPT's translations closely resemble NMT outputs rather than human translations. Supervised classifiers effectively differentiated the three translation types, while unsupervised clustering was less successful.

Benbada and Benaouda (2023) adapted a comparative analytical approach to investigate the role of AI in developing Machine translation quality. The sample of the study was a professional human translator and two different types of Machine Translation online are Google Translate, and Reverso. The study found limitations of machine translation, particularly in capturing contextual, idiomatic expressions, and culture-specific references.

Abdulaal (2022) studied machine and human translation errors in some literary texts with some implications for EFL translators. The study aimed to draw a comparison between some internet emerging applications used for machine translation (MT) and human translation (HT) in two of Alphonse Daudet's short stories. The automatic translation has been carried out by four MT online Applications including (Translation Dic, Yandex, Mem-Source, and Reverso). The findings of the study revealed that MT and HT made some errors related to polysemy, homonymy, syntactic ambiguities, fuzzy hedges, synonyms, metaphors and symbols.

Ali (2020) assessed the quality and machine translation by evaluating online machine translation of English into Arabic texts. The study compared the numbers and percentages of errors occurring in English into Arabic translation outputs using three MT applications. The method used in this research is a quantitative analysis of the number of errors related to the translation attributes. The sample of the study was machine translation (MT) outputs, an English text and its Arabic counterpart were selected from the UN records. The findings of this study imply that these MT applications can be implemented to perform English into Arabic translation to get the broad gist of a source text, but a deep and thorough post-editing process looks essential for a full and accurate understanding of an English into Arabic MT output.

Mudawe (2019) explored the potential of technology-based translation tools, including MT, Computer-Aided Translation (CAT), and Translation Management Systems (TMS). The study assessed Google Translate's performance compared to human translators, using Grammarly for quality evaluation. The findings of the study underscored the role of translation technology in bridging linguistic barriers while emphasizing the need for continued research to align automated translations with global standards.

The studies stated above were chronologically ordered; most of these studies focused on general translation quality and examined specific translation tools. They did not often provide comprehensive, tool-based analysis of the translation in media discourse. Thus, the gap that the present study seeks to address lies in an extensive comparative analysis of six AI translation tools across a diverse set of twenty Arabic news texts. By widely tackling linguistic and cultural challenges, this study provides a more detailed understanding of each tool's performance in the context of Arabic media translation. Further, the current study introduces a unique rubric incorporating a five-point scoring system to assess translation quality.

3. Study Methodology

3.1. Study Design: -

The study employed a quantitative research method to collect and analyze numerical data on translation errors, focusing on their frequency and distribution regarding their linguistic and formatting types. Meanwhile, this method was used to conduct a systematic evaluation of the performance of the selected AI translation tools, enabling a detailed assessment of their translation accuracy.

3.2. Corpus of the Study: -

The study's corpus comprises twenty Arabic news texts, sourced from a variety of reputable news websites, including Al-Jazeera Net, Russia Today, Al-Quds Al-Arabi, Asharq Al-Awsat, Marebpress, Alarabia Net, BBC Arabic. These texts were translated into English by the researchers, reviewed by experts, and used then as model translations for assessing the translation outputs generated by three AI-driven translation tools—Google Translate, Reverso, and Yandex, and three large language models—Chat GPT-4, Bing, and Gemini-1.5-Pro.

3.3. Data Collection: -

The data collection process began with the selection of the Arabic news texts from the aforementioned news websites. These texts were input into the six selected translation tools, and the resulting English translations were collected for further evaluation. The collected outputs were then classified, assessed, and analyzed in order to measure the performance of each tool and determine the levels of their translation accuracy.

3.4. Validation and Reliability: -

To ensure validity and reliability, the model translations were reviewed by three professors who are experts in English language and translation studies. These experts evaluated the translations for their clarity, linguistic accuracy, and contextual appropriateness, thereby validating their use as reliable benchmarks for assessing the tool-generated outputs.

3.5. Data Analysis Procedures: -

The study examined the outputs of translation tools through the following three-tiered analytical framework:

1- Error classification: translation errors were systematically categorized into three main types: lexico-semantic, syntactic, and formatting errors. Each category was then analyzed in detail to highlight the nature and frequency of the issues observed across the translated news texts.

2- Tools' Performance assessment: to evaluate the overall performance of the translation tools, the researchers developed a detailed evaluation rubric incorporating a five-point scoring system (8, 6, 4, 2, 0). Additionally, the model translations were used as benchmarks to ensure consistent and objective scoring. A clearly defined set of criteria, along with corresponding numerical ranges, was established to guide the assignment of scores based on how well each translation met the standards of accuracy and fidelity to the source text.

3- Determination of translation accuracy levels: the same rubric was employed to determine the levels of translation accuracy in the translated texts. The classification was conducted according to predefined criteria in the rubric, allowing the researchers to quantify and compare the accuracy levels across the different examined tools.

Table (1)

Level of Accuracy	Score	Numerical Range (out of 8)	Criterion description
Highly Accurate	8	7.5 - 8	The translation is flawless with no linguistic or formatting errors. It fully preserves the meaning, tone, and style of the target text. Terminology is precise and contextually appropriate, including media-specific terms. No or minimal post-editing is required.
Accurate	6	5.5 - 7.4	The translation is mostly correct with only minor linguistic or formatting errors that do not significantly impact comprehension. The meaning is well conveyed in the target text, though slight nuances due to missing. Terminology is mostly accurate, with only occasional misuse. Minor post-editing is required to improve clarity.
Moderately Accurate	4	3.5 - 5.4	The translation conveys the general meaning of the target text but contains noticeable linguistic or formatting errors. Some distortion or loss of meaning occurs in key phrases. Terminology is inconsistent, leading to occasional ambiguity.

			Moderate post-editing is required to correct errors and improve readability.
Less Accurate	2	1.5 - 3.4	The translation contains frequent major linguistic or formatting errors that significantly impact the meaning and readability of the target text. Key ideas are partially misrepresented, or unclear. Terminology is frequently incorrect, leading to misunderstandings. Extensive revision is required to correct errors and improve readability.
Inaccurate	0	0 - 1.4	The translation fails to convey the intended meaning of the target text. It contains severe linguistic and formatting errors. Terminology is incorrect or missing, leading to a completely misleading or unreadable translation. Complete retranslation is required.

Table (1) outlines the framework employed by three researchers to score translation outputs and determine levels of translation accuracy. Translation quality was assessed according to predefined criteria shown above.

4. Results and Discussion

4.1. Errors analysis: -

This part of the research focuses on the classification and analysis of errors made by the six translation tools when translating the twenty Arabic news into English. It answers the first question of the study: *What are the most common types of errors made by AI-driven translation tools: Google Translate, Reverso, Yandex, and large language models: Chat GPT-4, Bing, and Gemini-1.5-Pro when translating Arabic news texts into English?*

Table (2) includes the types of the errors found in the translated twenty Arabic news texts.

Text	Translation Tool	Error Types						Total
		Lexico-semantic Errors		Syntactic Errors		Formatting Errors		
		Frq.	PCT%	Frq.	PCT%	Frq.	PCT%	
Texts 1-7	Google Translate	15	29.41 %	6	24 %	3	8.57 %	24
	Reverso	14	27.45 %	10	40 %	6	17.14 %	30
	Yandex	8	15.69 %	5	20 %	13	37.14 %	26
	Chat GPT-4	3	5.88 %	1	4 %	0	0 %	4
	Bing	5	9.80 %	2	8 %	5	14.28 %	12
	Gemini-1.5-Pro	6	11.76 %	1	1 %	8	22.86 %	15
Texts 8-14	Google Translate	17	26.56 %	9	29 %	4	9.52 %	30
	Reverso	19	29.68 %	10	32.25 %	8	19 %	37
	Yandex	10	15.62 %	6	19.35 %	16	38.09 %	32
	Chat GPT-4	4	6.25 %	2	6.45 %	0	0 %	6
	Bing	6	9.37 %	2	6.45 %	7	16.66 %	15
	Gemini-1.5-Pro	8	12.5 %	2	6.45 %	7	16.66 %	17
Texts 15-20	Google Translate	28	28.57 %	12	24 %	6	8 %	46
	Reverso	28	28.57 %	20	40 %	13	17.33 %	61
	Yandex	15	15.30 %	9	18 %	24	32 %	48
	Chat GPT-4	6	6.12 %	3	6 %	0	0 %	9
	Bing	9	9.18 %	3	6 %	14	18.67 %	26
	Gemini-1.5-Pro	12	12.24 %	3	6 %	18	24 %	33
	Total	213	45.22 %	106	22.50 %	152	32.27 %	471

The previous table shows the categorization, frequency, and percentage of errors in the translated twenty Arabic news texts into English. These errors are classified into three main categories: lexico-semantic, syntactic and formatting errors. Additionally, the texts are divided into three groups for analysis; the first group comprises texts one to seven, the second group includes texts eight to fourteen, and the third group consists of texts sixteen to twenty. This classification facilitates the analysis of calculated data and error distribution for all texts across the six translation tools in the next table.

4.2. Analysis of Error Types: -

Table (3) includes a summary of error types for twenty news texts made by the six translation tools.

Text	Translation Tool	Error Types						Total	PCT%
		Lexico-semantic Errors		Syntactic Errors		Formatting Errors			
		Frq.	PCT%	Frq.	PCT%	Frq.	PCT%		
Text 1-20	Google Translate	60	28.17 %	27	25.47 %	13	8.55 %	100	21.23 %
	Reverso	61	28.63 %	40	37.73 %	27	17.76 %	128	27.19 %
	Yandex	33	15.5 %	20	18.86 %	53	34.86 %	106	22.50 %
	Chat GPT-4	13	6.10 %	6	5.66 %	0	———	19	4.03 %
	Bing	20	9.38 %	7	6.60 %	26	17.10 %	53	11.25 %
	Gemini-1.5-Pro	26	12.20 %	6	5.66 %	33	21.71 %	65	13.80 %
	Total	213	45.22 %	106	22.50 %	152	32.27%	471	100%

Based on the data shown in Table (3) above, it seems that a total of (471) linguistic errors were identified across all translation tools when translating twenty news texts from Arabic into English. These errors are classified into three main categories: lexico-semantic errors, syntactic errors, and formatting errors. The higher number of these errors is found in the lexico-semantic category, with 213 errors, accounting for 45.22 % of the total errors, followed by the formatting errors with 152 instances, representing 32.27% of the overall errors. The syntactic errors category has the lowest occurrence, with 106 errors, comprising 22.50 % of the total errors. A detailed analysis and interpretation of these errors are provided in the following sections.

4.2.1. Lexico-semantic Errors.

The range of (213) lexico-semantic errors varies among the six translation tools, indicating significant variations in their ability to handle word meanings and contextual accuracy. These differences suggest that some tools rely heavily on literal, word-for-word translations, while others demonstrate a better grasp of contextual adaptation. Notably, Reverso and Google Translate stand out as the most error-prone tools in this category, making 61 errors (28.63%) and 60 errors (28.17%), respectively. Their high error rates highlight their struggle with selecting the correct words in context, often leading to inaccurate translations. This issue may arise from their algorithmic approach, which prioritizes direct lexical correspondences rather than interpreting meaning based on surrounding textual clues. Such a pattern reveals a fundamental limitation in their ability to handle nuanced expressions and polysemous words, which are critical in media translation. Yandex, which made 33 lexico-semantic errors (15.5%), performed slightly better than Google Translate and Reverso but still showed notable challenges in maintaining precise word meanings. While it exhibited some contextual awareness, it remained prone to misinterpretations, particularly in complex sentences where word choices require deeper semantic understanding.

Gemini-1.5-Pro, on the other hand, recorded 26 errors (12.20%), demonstrating a moderate improvement in lexico-semantic accuracy. Although it still struggled with word choice in certain cases, its overall performance reflected a better ability to process meaning within context compared to more literal-based tools. This indicates that Gemini-1.5-Pro is somewhat more reliable in understanding the relationship between words in a sentence, though occasional inconsistencies persist.

Bing, with 20 errors (9.38%), showed a noticeable reduction in lexico-semantic errors, suggesting a stronger capability to preserve meaning beyond individual word translation. This indicates a greater

ability to interpret phrases contextually rather than relying solely on direct translations, making it a more effective tool for retaining coherence in media texts. However, the most notable performance in this category comes from Chat GPT-4, which made only 13 lexico-semantic errors (6.10%), the lowest among all tools. This result suggests that Chat GPT-4 excels in word choice accuracy and contextual adaptation, reflecting a superior understanding of meaning nuances. Its ability to process idiomatic expressions, polysemous words, and domain-specific terminology more effectively than other tools; this result makes it the most reliable option for lexico-semantic accuracy in media translation.

4.2.2. Syntactic Errors.

The 106 syntactic errors recorded across the six translation tools reveal notable differences in their ability to handle sentence structure, grammar, and syntax rules. Among them, Reverso exhibited the highest number of syntactic errors, with 40 errors (37.73%) while translating the twenty news texts, indicating significant struggles in maintaining the proper usage of the verb form, and grammatical agreement. This suggests that Reverso often fails to adapt sentence structures effectively when translating between languages, which can lead to unnatural or grammatically incorrect sentences.

Google Translate, with 27 errors (25.47%), performed better than Reverso but still displayed substantial syntactic weaknesses. These errors likely tend to arise from a literal translation without fully adjusting to the target language's grammatical structure, leading to problems such as incorrect verb form and article usage, awkward phrasing and incorrect sentence formations.

Yandex, with 20 errors (18.86%), demonstrated a more moderate level of syntactic accuracy, performing better than both Reverso and Google Translate. However, its errors indicate ongoing difficulties with sentence construction, particularly in misusing of definite articles or more complex sentences where word order and clause relationships are critical. While Yandex shows some level of syntactic awareness, it remains prone to errors in maintaining proper grammatical coherence. Bing, on the other hand, made only 7 syntactic errors (6.60%), a significant reduction compared to the previous tools. This suggests that Bing has a stronger grasp of sentence structure and grammar, allowing it to produce more coherent and grammatically sound translations. Its ability to maintain sentence integrity indicates a better understanding of syntax compared to the more error-prone tools.

Chat GPT-4 and Gemini-1.5-Pro recorded the lowest number of syntactic errors, with only 6 errors each (5.66%). This suggests that these two tools exhibit the highest syntactic accuracy among all tested translation tools. Their low error rates imply a strong ability to adapt sentence structures appropriately, ensuring grammatical accuracy and fluency in the translated text. The minimal syntactic errors made by Chat GPT-4 and Gemini-1.5-Pro highlight their effectiveness in maintaining correct word order, subject-verb agreement, and sentence coherence. Their performance suggests that they are the most reliable options for translations requiring syntactic precision, making them well-suited for translating complex sentences while preserving grammatical integrity.

4.2.3. Formatting Errors.

The 152 formatting errors observed across the six translation tools highlight significant differences in their ability to preserve text structure, misplaced punctuation, spacing, and irregular capitalization. Among them, Yandex exhibited the highest number of formatting errors, with 53 errors (34.86%), indicating a major struggle in maintaining sentence layouts and structural elements. This suggests that Yandex frequently alters text formatting by disrupting spacing, capitalization, punctuation, and line breaks, making it less reliable for translating structured media content where proper formatting is essential for readability and clarity. Gemini-1.5-Pro also displayed a considerable number of formatting issues, with 33 errors (21.71%), suggesting that it, too, has difficulties in handling structured text elements. However, compared to Yandex, Gemini-1.5-Pro demonstrated slightly better consistency in formatting, indicating that while it still struggles, it does not distort text structure as severely.

Reverso made 27 formatting errors (17.76%), placing it in the middle range of performance. While it showed a moderate level of formatting accuracy, it still struggles with maintaining structured text, capitalization and punctuation. This suggests that Reverso, while somewhat better than Yandex and Gemini-1.5-Pro, still alters text layout to a noticeable extent. Bing, with 26 formatting errors (17.10%), performed similarly to Reverso, displaying moderate challenges in maintaining text format. The errors

seen in Bing's translations involve minor inconsistencies in punctuation, spacing, or text alignment, which, while not as severe as Yandex's, still impact the final presentation of translated media content. Google Translate, with only 13 formatting errors (8.55%), performed significantly better in preserving text structure compared to the other tools. Its relatively low formatting error count suggests that it is more reliable in maintaining layout elements such as spacing, punctuation, and text segmentation. However, occasional inconsistencies remain, particularly in punctuation placement and spacing adjustments. Notably, Chat GPT-4 made no formatting errors, making it the most reliable tool in terms of structural preservation. This suggests that Chat GPT-4 excels at maintaining the original layout, ensuring that line breaks, punctuation, and text alignment remain intact, making it the ideal choice for translating structured news texts where formatting consistency is crucial.

4.3. Error Frequency of Translation Tools:

Based on the data presented in Table (3), this section addresses the first part of the second research question: *Which of the investigated translation tools exhibits the lowest frequency of errors in translating Arabic news texts into English?*

The analysis of each translation tool reveals notable differences in performance and reliability. Reverso made the highest number of errors across the twenty news texts, totalling 128 errors out of 471, which represents 27.19% of all recorded errors. These errors were distributed across 61 lexico-semantic, 40 syntactic, and 27 formatting errors. Lexico-semantic errors were particularly evident, such as the mistranslation of the phrase "مواطنين ينوحون على أهاليهم", *muwāṭinīn yanūḥūn ʿalā ahālīhim*, which Reverso translated it as "citizens pay tribute to their parents". This translation significantly distorts the intended meaning "citizens mourn their relatives," substituting emotional mourning with admiration. It also incorrectly translated "أهاليهم" *ahālīhim* as "parents". Such errors reflect a reliance on literal, word-for-word translation rather than contextual interpretation. Syntactic errors made by Reverso included incorrect verb forms and tense usage. For instance, the phrase "لن تحضرا" *lan taḥḍurā* was translated as "would not attend" instead of the correct "will not attend," thus altering the intended future tense. Reverso also exhibited formatting issues, such as inconsistent punctuation—failing to replicate a full stop in the phrase "في سوريا" *fī Sūriyā*, instead using a comma, "in Syria,"—as well as lexical errors like rendering "عشرات الآلاف" *asharāt al-ālāf* as "Tens of 1000" rather than "Tens of thousands." Yandex ranked second in error frequency, making 106 errors (22.50%), distributed as 33 lexico-semantic, 20 syntactic, and 53 formatting errors. One notable example of a lexico-semantic error is its translation of the Arabic verb "تدفق" *tadaffuq* as "poured into," a phrase that typically conveys liquid movement or financial investment rather than the intended meaning of "converged on." Another example is the translation of "ميدان" *maydān* as "Maidan" rather than translating it correctly as *"square,"* this indicates a failure in cultural and contextual adaptation. Formatting errors were the most frequent in Yandex's outputs, including improper noun capitalization—for example, the translation of "اللاجئين" *al-lājiʾīn* as "Refugees", was capitalized mid-sentence instead of the correct form "refugees". In terms of syntactic, Yandex exhibited frequent errors such as omitting definite articles by translating " الحرب اليمنية" *al-ḥarb al-yamanīyah* as "the Yemeni war" without "the." And in handling incorrect verb forms, such as translating "لمنع مقتل المزيد من المدنيين" *limanʿ maqtal al-mazīd min al-madanīyīn* as " prevent the death of more civilians" instead of " to prevent/ or preventing the death of more civilians". These issues reflect a lack of grammatical consistency and misapplication of English article usage.

Google Translate, on the other hand, accounted for 100 errors (21.23%), consisting of 60 lexico-semantic, 27 syntactic, and 13 formatting errors. Notable lexico-semantic errors include mistranslating many Arabic words into English, for instance, the Arabic phrase *"قمتين عربية وخليجية" qimmtayn ʿArabīyah Khalījīyah* was translated as "Arab and Gulf games", whereas the correct translation should be "Arab and Gulf summits.". This misinterpretation not only distorts the intended meaning but also misleads the reader, especially in politically sensitive contexts where the difference between games and summits is significant. Another example is the misinterpretation of the Arabic word *"اعترف" iʿtarafa*, which was incorrectly translated as "said" rather than the more accurate and contextually appropriate "admitted." This reflects a lack of semantic nuance in distinguishing between the meaning differentiations. Syntactically, Google Translate struggled with verb tenses and determiner usage. An example is the translation of "زار اليوم" zāra al-yawm as "visited today." Although grammatically correct, a more contextually appropriate rendering in English media discourse would be "has visited

today," better aligning with journalistic conventions. Additional errors were observed in determiner usage, particularly in Google Translate outputs. For instance, the phrase "للمطالبة بحكم مدني" *lil-mutālabah bi-ḥukm madanī* was translated as "to demand a civilian rule", where the use of the indefinite article "a" is grammatically inappropriate. In standard English usage, "civilian rule" typically appears without an article, making the translation both awkward and incorrect. Although formatting errors were less frequent in Google Translate, they still impacted overall translation quality. These included inconsistent punctuation, incorrect capitalization—such as translating "صواريخ" *ṣawārīkh* as "Rockets" mid-sentence—and omissions of diacritical markers like apostrophes, as seen in "صنعاء" *Ṣan‘ā’* rendered as "Sanaa" instead of "Sana'a." These examples highlight that while Google Translate often preserves basic structural integrity, it still struggles with contextual meaning, stylistic nuance, and domain-appropriate language, particularly in political and media-related texts.

Gemini-1.5-Pro, meanwhile, recorded a total of 65 errors (13.80%), including 26 lexico-semantic, 6 syntactic, and 33 formatting errors. In the lexico-semantic category, Gemini-1.5-Pro frequently opted for literal translations that were contextually inaccurate. For example, it translated"القوى السياسية" *al-quwā al-siyāsiyyah* as "political forces" rather than the more appropriate "political parties," showing a lack of sensitivity to domain-specific terminology. Similarly, "زرع الانقسام في المنطقة" *zar‘ al-inqisām fī al-mintaqah*, was translated as "sowing division," a phrase that implies agricultural activity, whereas "deepening divisions" would have more accurately captured the intended figurative meaning. Syntactic errors included occasional tense mismatches, such as in the translation of "لم يتم بعد تحديد الموعد النهائي" *lam yatimm ba‘d taḥdīd al-maw‘id al-nihā’ī* as "the time of talks had not yet been set," which incorrectly uses the past perfect tense instead of the more accurate present perfect form, "has not yet been set". Formatting issues were also prominent in Gemini-1.5-Pro’s output. These included misplaced bullet points and inconsistent capitalization—for example, rendering "دول الخليج" *duwal al-khalīj* as "Gulf states" instead of "Gulf States," and "الخرطوم" *al-Kharṭūm* as "khartoum" instead of "Khartoum." While Gemini-1.5-Pro showed relative strength in syntactic consistency, its weaknesses in lexical precision and formatting reduced the overall reliability of its translations.

Bing made 53 errors (11.25%), broken down into 20 lexico-semantic, 7 syntactic, and 26 formatting errors. Lexico-semantic errors included the imprecise translation of culturally and contextually loaded terms. For example, the Arabic phrase "فرض علينا" *faraḍa ‘alaynā* was translated as "imposed on us," which carries a coercive tone not intended in the original religious context; a more accurate translation would be "an obligation upon us." Syntactic issues were primarily related to article usage. Bing occasionally inserted definite articles where they were unnecessary, such as translating "في سجون" *fī sujūn* as "in the prisons", which grammatically is used without "the", instead of the correct "in prisons." Formatting errors included inappropriate capitalization, such as "On the occasion" for "بمناسبة الذكرى," *bimunāsabat al-dhikrā* where "on" was incorrectly capitalized mid-sentence. Despite these issues, Bing demonstrated comparatively strong grammatical coherence and maintained a more accurate sentence structure than most tools analysed.

Chat GPT-4, by contrast, recorded the lowest number of errors across all tools, with only 19 errors (4.03%), comprising 13 lexico-semantic and 6 syntactic errors, and no formatting errors. Among the lexico-semantic errors were minor word choice issues—for example, translating "عدد كبير من المصابين" *adad kabīr minal-muṣābīn* as "big number of casualties"; the adjective "big" is semantically inappropriate. The more accurate and contextually suitable choice is "large number of casualties". Additionally, Chat GPT-4 translated the Arabic term "جمود" in the phrase "جمود المفاوضات" *jumūd al-mufāwaḍāt* as "freezing of talks" rather than the idiomatic "deadlock in negotiations." Syntactically, Chat GPT-4 occasionally misused definite and indefinite articles, such as translating "للمطالبة بحكم مدني" *lil-mutālabah bi-ḥukm madanī* as "a civilian rule," where the article "a" should have been omitted, and rendering *al-ḥarb al-yamanīyah* "الحرب اليمنية" as "Yemeni war" without the definite article "the". Despite these minor errors, Chat GPT-4’s complete accuracy in formatting, combined with its minimal lexico-semantic and syntactic issues, indicates that it is the most reliable translation tool among those evaluated. Its ability to preserve meaning, maintain grammatical coherence, and adhere to stylistic conventions makes it particularly effective for translating Arabic news into English.

4.4. Assessment of the Translation Tools Performance: -

This section addresses the second part of the second research question: *Which of the investigated translation tools demonstrate better performance in translating Arabic news texts into English?* The researchers employed the aforementioned graded scoring rubric to assess the tools' performance and the quality of each translation. Additionally, the model translations were used as benchmarks to ensure consistency and objectivity in scoring. Each researcher independently assessed and scored the translation outputs using a five-point scale (8, 6, 4, 2, 0), based on how closely each output aligned with the rubric's accuracy and model translations.

Table (4)

Text No.	Google Translate			Reverso			Yandex			Chat GPT-4			Bing			Gemini-1.5-Pro		
	Researchers			Researchers			Researchers			Researchers			Researchers			Researchers		
	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3	R1	R2	R3
Text 1	7	6	7	6	5	6	5	6	6	8	8	8	6	6	7	7	7	7
Text 2	6	5	6	5	6	6	7	7	6	8	7	8	7	6	7	7	7	7
Text 3	6	7	7	4	5	6	5	5	5	7	8	8	8	7	8	7	6	7
Text 4	5	6	6	6	5	5	6	6	7	7	8	7	7	7	8	6	6	6
Text 5	6	6	7	7	7	7	7	7	6	8	8	8	7	7	8	7	7	7
Text 6	6	7	6	5	6	6	6	5	7	7	7	8	6	8	7	6	7	7
Text 7	7	6	6	7	6	6	7	7	7	8	8	8	8	8	8	7	7	7
Text 8	6	6	7	5	6	6	6	6	6	7	7	8	7	6	8	8	7	8
Text 9	5	5	6	6	6	5	6	6	6	8	8	7	7	7	7	6	7	6
Text 10	6	7	7	6	6	6	7	6	7	7	7	8	8	7	7	7	8	7
Text 11	7	7	6	6	6	7	7	7	6	7	8	8	7	8	8	8	7	7
Text 12	6	6	6	5	5	6	5	7	6	8	7	8	8	8	6	8	7	6
Text 13	6	7	6	6	6	7	7	6	5	8	8	7	6	7	7	6	7	7
Text 14	7	6	6	5	6	6	5	6	6	8	7	8	7	6	8	7	6	7
Text 15	7	6	7	6	6	6	6	6	6	8	8	8	7	7	7	7	6	7
Text 16	6	6	7	5	7	5	7	7	6	8	7	7	8	7	7	6	7	7
Text 17	6	7	7	6	6	7	7	7	7	8	7	8	7	8	7	7	7	8
Text 18	5	6	6	6	5	5	5	6	6	7	7	8	7	7	6	6	7	7
Text 19	7	6	7	6	7	6	6	7	6	8	8	8	6	7	7	7	7	7
Text 20	6	6	6	5	6	5	7	6	5	8	7	8	7	8	7	7	6	7

Table (4) above presents the compiled scores assigned by the researchers of the study to the translations generated by the six translation tools.

Table (5)

Text No.	Google Translate	Reverso	Yandex	Chat GPT-4	Bing	Gemini-1.5-Pro
	Mean Scores	Mean Scores	Mean Scores	Mean Scores	Mean Scores	Mean Scores
Text 1	6.66	5.66	5.66	8	6.33	7
Text 2	5.66	5.66	6.66	7.66	6.66	7
Text 3	6.66	5	5	7.66	7.66	6.66
Text 4	5.66	5.33	6.33	7.33	7.33	6
Text 5	6.33	7	6.66	8	7.33	7
Text 6	6.33	5.66	6	7.33	7	6.66
Text 7	6.33	6.33	7	8	8	7
Text 8	6.33	5.66	6	7.33	7	6.66
Text 9	5.33	5.66	6	7.66	7	6.33
Text 10	6.66	6	6.66	7.33	7.33	7.33
Text 11	6.66	5.66	6.66	7.66	7.66	7.33
Text 12	6	5.33	6	7.66	7.33	7
Text 13	6.33	6.33	6	7.66	6.66	6.66
Text 14	6.33	5.66	5.66	7.66	7	6.66
Text 15	6.66	6	6	8	7	6.66
Text 16	6.33	5.66	6.66	7.33	7.33	6.66
Text 17	6.66	6.33	7	7.66	7.33	7.33
Text 18	5.66	5.33	5.66	7.33	6.66	6.66
Text 19	6.66	6.33	6.33	8	6.66	7
Text 20	6	5.33	6	7.66	7.33	6.66
Total means	6.26	5.94	6.17	7.68	7.08	6.86

Table (5) presents a summary of the mean scores assigned by the researchers for 20 texts across the six translation tools. The calculation of these means was conducted using the following formula:

Mean= $\frac{Researcher\ 1+Researcher\ 2+Researcher\ 3}{Three}$ =

Table (6)

Translation Tool	Score Average	Percentage
Google Translate	6.26	78.25 %
Reverso	5.94	74.25 %
Yandex	6.17	77.12 %
Chat GPT	7.68	96 %
Bing	7.08	88.5%
Gemini-1.5-Pro	6.86	85.75 %

Table (6) provides an average summary of the evaluation results. To calculate the overall mean scores for the translation outputs of news texts, the researchers applied the following formula: (Total mean = $\frac{\sum Mean}{Total(20)}$ =). Additionally, the percentage was calculated using the formula: (Total mean = $\frac{\sum Mean}{Total\ (8)}$ =).

Based on the summary shown in Table (6), the mean scores assigned by the three researchers for the translation of the twenty Arabic news texts reveals a clear range in the performance of the evaluated translation tools. The average scores vary between 7.68 (Chat GPT-4) and 5.94 (Reverso), with the other tools falling in between this range. Chat GPT-4 achieved the highest average score of 7.68 out of 8 (96%). This indicates exceptional performance in translating the content of the twenty news texts with a highly accurate. The translated texts exhibit a high level of lexical appropriateness, grammatical precision, and formatting accuracy, with only minimal errors observed compared to the expert-reviewed translation. Bing ranked second with a score of 7.08 out of 8 with a percentage of 88.5%. This suggests that the performance of this AI tool is strong as it accurately translated the texts from Arabic into English. It only made fewer lexical, syntactic, or formatting errors compared to the human-reviewed translations. Gemini-1.5-Pro tool received a slightly higher average score of 6.86 points with a percentage of 85.75% for its translation of the twenty texts. While still achieving a reasonably high score, it indicates that there are only a few lexical, syntactic, and formatting errors compared to the top-performing tools.

Google Translate obtained a good score of 6.26 out of 8 (78.25%) for its translation of the twenty news texts. This suggests that Google Translate succeeded in conveying much of the intended meaning, correct grammatical structure, and formatting layout. Similarly, Yandex scored 6.17 out of 8 (77.12 %), placing it in the same performance tier. This suggests that it also did well as it exhibits some lexical, syntactic, and formatting errors for media texts. It is comparatively lower than the other tools evaluated by the researchers in the study. Reverso received the lowest average score of 5.94 out of 8, (74.25%). This result indicates that its translation contained a comparatively higher number of lexical, syntactic, and formatting errors. These issues frequently obscured or altered the meaning of the source text, making Reverso the least effective tool among those evaluated in the study.

Overall, the comparative analysis of these tools reveals a clear performance hierarchy. Chat GPT-4 appears the top performer and the most accurate and reliable, followed by Bing and Gemini-1.5-Pro, both of which demonstrated strong and consistent performance. Google Translate and Yandex show fair accuracy but are more prone to errors in complex linguistic structures. Reverso falls short in several areas and requires significant improvements to match the quality of the other tools.

4.5. Evaluation of Levels of Translation Accuracy: -

This part answers the third question of the study: *To what extent does translation accuracy vary among the studied tools when translating Arabic news texts into English?*

Based on the developed rubric, the researchers categorized the translation accuracy of each tool into five distinct levels: highly accurate, accurate, moderately accurate, less accurate, and inaccurate. These classifications were based on the overage scores assigned to the translation texts, as detailed in Table (6). The score range for each accuracy level was defined as follows:

Highly accurate (7.5- 8), accurate (5.5- 7.4), moderately accurate (5.5- 5.4), less accurate (1.5- 3.4), and inaccurate (0- 1.4).

Table (7) presents the classification of translation accuracy levels, along with the corresponding frequency and percentage values for each category, based on the translations of twenty Arabic news texts generated by the three AI-driven translation tools and the three large language models.

Translation Tool	Highly Accurate Translation		Accurate Translation		Moderately Accurate Translation		Less Accurate Translation		Inaccurate Translation	
	Frq.	PCT%	Frq.	PCT%	Frq.	PCT%	Frq	PCT%	Frq.	PCT %
Google Translate	0	——	19	95%	1	5%	0	——	0	——
Reverso	0	——	15	75%	5	25%	0	——	0	——
Yandex	0	——	19	95%	1	5%	0	——	0	——
Chat GPT-4	15	75 %	5	25%	0	——	0	——	0	——
Bing	3	15 %	17	85%	0	——	0	——	0	——
Gemini-1.5-Pro	0	——	20	100%	0	——	0	——	0	——

The table above summarizes the levels of translation accuracy achieved by the six translation tools when translating a set of twenty Arabic news texts into English. The evaluation framework categorized the translation accuracy into five distinct levels: highly accurate, accurate, moderately accurate, less accurate, and inaccurate. Notably, the results show that all translated outputs fell within the first three levels only— no text was evaluated as either less accurate or inaccurate, which highlights the relatively high baseline performance of all the tools examined in the study.

Among the tools, Chat GPT-4 achieved the highest level of accuracy, with 15 texts out of 20 texts rated as highly accurate, and the remaining 5 texts were accurately translated. This outstanding performance indicates that Chat GPT-4 consistently preserved the meaning, structure, and style of the original Arabic media content, with minimal linguistic or formatting deviations. The high proportion of texts rated in the top tier reflects Chat GPT-4's advanced capabilities in managing complex sentence structures, media terminology, and idiomatic expressions. Bing also showed a strong performance. It produced 3 highly accurate translations, while 17 texts were assessed as accurate. This result places Bing second in terms of overall performance. Although it achieved fewer highly accurate translations than Chat GPT-4, the consistency in accurate output suggests that Bing is reliable and competent, handling most of the texts with a good degree of linguistic and contextual fidelity.

In contrast, all AI-driven translation tools: Google translate, Reverso, and Yandex failed to highly accurately translate any news texts from Arabic into English. For example, Gemini-1.5-Pro translated all 20 news texts with an accurate rating. While this consistency demonstrates a solid baseline of translation quality, it also indicates that the tool struggles to achieve excellence in capturing finer nuances or higher syntactic fluency that would elevate a translation to the highest tier. Nonetheless, its uniform performance suggests that it is somewhat suitable for contexts requiring reliable, general-purpose translations. Moreover, both Google Translate and Yandex showed similar performance patterns. Each tool translated 19 out of 20 texts accurately, with 1 text classified as moderately accurate. This result reflects a generally good ability to convey meaning but indicates occasional issues, possibly related to idiomatic or context-sensitive language. The absence of highly accurate outputs from both tools suggests that while they are effective, their translations often require revision to meet professional or academic standards. Finally, Reverso demonstrated the lowest performance among the tools in this category. It translated 15 out of 20 texts accurately and 5 texts with moderate accuracy. This distribution indicates that while Reverso performed reasonably well in the majority of cases, it faced greater challenges than the other tools in maintaining consistency and fluency. The presence of five moderately accurate translations implies that Reverso requires more substantial human post-editing when used for formal or high-precision translation tasks.

1. Conclusion

The present study aimed to evaluate the performance and translation accuracy of AI-driven translation tools and large language models in translating selected twenty Arabic news texts into English. The study selected three AI-driven translation tools —Google Translate, Reverso, Yandex, — and three large language models: Chat GPT-4, Gemini-1.5-Pro and Bing.
To ensure a systematic evaluation, translation errors were classified into three main categories: lexico-semantic, syntactic, and formatting errors. The analysis revealed significant variations in performance across the tools. Lexico-semantic accounted for the highest number of errors, highlighting persistent challenges in word choice and contextual accuracy. Formatting errors were the second highest errors, showing clear difficulties in preserving text layout and structure. Syntactic errors were classified third, reflecting difficulties in adapting grammatical and sentence structures between the two languages.

Among all the tools assessed, Chat GPT-4 exhibited the highest level of accuracy, producing the fewest errors across all categories. In contrast, Reverso recorded the highest number of errors. Regarding the performance evaluation of these six translation tools, the score averages across the different translation tools show that Chat GPT-4 achieved the highest average score among all tools, significantly outperforming the others. In contrast, Reverso received the lowest score, indicating the weakest performance. The remaining tools demonstrated moderate results, with Yandex and Google yielding comparable scores, and Bing and Gemini-1.5-Pro also performing at relatively similar levels.

In terms of translation accuracy, ChatGPT-4 again ranked the highest, demonstrating superior capability in translating the selected Arabic news texts into English. Bing followed, albeit with a noticeable performance gap. The other tools—Google Translate, Reverso, Yandex, and Gemini-1.5-Pro—struggled to produce highly accurate translations of the selected news texts, suggesting that such texts remain a significant challenge for most studied tools. These findings indicate that large language models outperformed the AI-driven translation tools in both accuracy and performance, particularly Chat GPT-4, representing a more reliable option for translating Arabic news content into English.

Additionally, the study suggested the importance of selecting appropriate translation tools based on the required level of accuracy, especially in critical fields such as media, academia, and formal communication, where linguistic precision and cultural sensitivity are essential. Although large language models represent a promising advancement in the field of machine translation, their outputs still require human post-editing to ensure the highest level of accuracy and quality.

Future research is recommended to explore translation performance across other specialized domains, such as legal, literary, and religious texts, using a broader range of tools. Additionally, refining machine translation evaluation metrics and leveraging large-scale, domain-specific datasets could contribute to the continued improvement of AI-driven translation tools and large language models, particularly in Arabic-language contexts.

References

Abdelaal, N., & Alazzawie, A. (2020). Machine translation: The case of Arabic-English translation of news texts. *Theory and Practice in Language Studies, 10*(4), 408–418.

Abdulaal, M. A. A.-D. (2022). Tracing machine and human translation errors in some literary texts with some implications for EFL translators. *Journal of Language and Linguistic Studies, 18,* xx–xx.

Ahmed, M. N. (2024). Ethics of translation and journalism: Truth, accuracy and cultural sensitivity in media communication. *Al-Noor Journal for Digital Media Studies, 1*(3), 35–45.

Ali, M. A. (2020). Quality and machine translation: An evaluation of online machine translation of English into Arabic texts. *Open Journal of Modern Linguistics, 10*(5), 524–548.

Al-Maaytah, M., & Almahasees, Z. (n.d.). A linguistic investigation for a case study of ChatGPT and Google Translate in rendering special needs texts from English into Arabic: A synchronic case study. *Journal name*, volume(issue), pages.

Almaaytah, S. A., & Alzobidy, S. A. (2023). Challenges in rendering Arabic text to English using machine translation: A systematic literature review. *IEEE Access, 11,* 94772–94779. https://doi.org/xxxxx

Al-Salman, S., & Haider, A. S. (2024). Assessing the accuracy of MT and AI tools in translating humanities or social sciences Arabic research titles into English: Evidence from Google Translate, Gemini, and ChatGPT. *International Journal of Data and Network Science, 8*(4), 2483–2498.

Benbada, M. L., & Benaouda, N. (2023). Investigation of the role of artificial intelligence in developing machine translation quality: Case study of Reverso Context and Google Translate translations of expressive and descriptive texts (Arabic-English/English-Arabic) (Doctoral dissertation, Faculty of Letters and Languages–Department of English).

Chacha, L., & Mwangi, I. (2024). Challenges of translating conversational implicatures from English to Kiswahili using computer-assisted tools: A case of Google Translate. *Mwanga wa Lugha, 9*(1), 129–135.

Chandra, R., Chaudhary, A., & Rayavarapu, Y. (2025). An evaluation of LLMs and Google Translate for translation of selected Indian languages via sentiment and semantic analyses. *arXiv preprint arXiv:2503.21393.*

Chen, L., Wang, W., & Hu, D. (2024). E³: Optimizing language model training for translation via enhancing efficiency and effectiveness. In *China National Conference on Chinese Computational Linguistics* (pp. 75–90). Singapore: Springer Nature Singapore.

Deng, L. (2016). Deep learning: From speech recognition to language and multimodal processing. *APSIPA Transactions on Signal and Information Processing, 5,* e1.

Falempin, A., & Ranadireksa, D. (2024). Human vs. machine: The future of translation in an AI-driven world. In *Widyatama International Conference on Engineering 2024 (WICOENG 2024)* (pp. 177–183). Atlantis Press.

Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: Human translation vs. AI large language models (Gemini and ChatGPT). *Cogent Social Sciences, 10*(1), 2410998.

Jiang, Z., Lv, Q., Zhang, Z., & Lei, L. (2023). Distinguishing translations by human, nmt, and chatgpt: A linguistic and statistical approach. *arXiv preprint arXiv:2312.10750*.

Mohsen, M. (2024). Artificial intelligence in academic translation: A comparative study of large language models and Google Translate. *Psycholinguistics, 35*(2), 134–156.

Mudawe, O. (2019). Ramping the Future of Translation Studies through Technology-based Translation. *International Journal of Comparative Literature and Translation Studies*, *7*(3), 74. https://d1wqtxts1xzle7.cloudfront.net/123295730/4097-libre.pdf

Ravshanovna (2024). *The role of technology in translation: From CAT tools to AI-driven translation*. *Modern Educational System and Innovative Teaching Solutions, 1*(4). https://esiconf.org/index.php/MESAS/article/view/1270/1187

fia, S. S. H. (2021). News and news translation: History and strategies. *Turkish Journal of Computer and Mathematics Education, 12*(11), 6710–6719.

Sholikhah, N. F. M., & Indah, R. N. (2021). Common lexical errors made by machine translation on cultural text. *Edulingua: Jurnal Linguistik Terapan dan Pendidikan Bahasa Inggris, 8*(1), 39–50.

Sidiya, A. M., Alzaher, H., Almahdi, R., & Elkafrawy, P. (2024). From analysis to implementation: A comprehensive review for advancing Arabic-English machine translation. In *2024 21st Learning and Technology Conference (L&T)* (pp. 109–114). IEEE.

Siu, S. C. (2024). Revolutionising translation with AI: Unravelling neural machine translation and generative pre-trained large language models. In *New advances in translation technology: Applications and pedagogy* (pp. 29–54). Singapore: Springer Nature Singapore.

Tekgurler, M. (2025). LLMs for translation: Historical, low-resourced languages and contemporary AI models. *arXiv preprint arXiv:2503.11898.*

Zanaty, D. G. (2024). When translating from Arabic to English and vice versa, is Google Bard a trustworthy tool? *مجلة وادي النيل للدراسات والبحوث الإنسانية والاجتماعية والتربوية، ٤٤*(44), 205–236.

Zinhom, H. (2024). The challenges of using machine translation in rendering Arabic texts into English: Applied perspective. *Journal for Foreign Languages, 16*(1), 175–198.