



Generative AI Potentials and Limits in Translating Iraqi Dialect in Literature: An Evaluation of Accuracy in Azher Jirjees's *At Rest in the Cherry Orchard*

Lect. Dr. Ruqaya Sabeeh Khalaf Al-Taie

Dept. of Translation/ College of Arts/ University of Basrah

Email: Ruqaya.sabeeh@uobasrah.edu.iq

Received May.3, 2026

Revised May.28, 2026

Accepted Jun.3, 2026

Online October1, 2026

ABSTRACT

Recent academic scrutiny has considered the usage of artificial intelligence (AI) tools in transforming literary texts and studies have largely shown that AI translation tools can generate accurate outputs for common vocabulary. However, their performance may vary when dealing with difficult expressions related to dialects, cultural distinctions, and context-dependent meanings. In literary translation from Arabic, maintaining a high level of accuracy in translating cultural expressions is challenging for AI systems. The main objective of this paper is to evaluate the accuracy level of four AI translations (ChatGPT, Microsoft Copilot, DeepSeek, and Gemini) and one human translation of Iraqi dialect in a novel by the Iraqi writer Azher Jirjees' *At Rest in the Cherry Orchard* which has been translated by Jonathan Wright into English. In the current study, the researcher will not only assess the level of accuracy of translation, but also evaluate the similarities and differences between the human translation and AI translations of the Iraqi dialect to see which is more reliable in understanding it. The researcher has adopted Nababan's translation accuracy theory to analyse the data in terms of accurate, less accurate, and inaccurate renditions. The analysis has revealed that the level of accuracy is equal in the human translation, DeepSeek, and Gemini at 67% and 16% inaccuracy, while Copilot and ChatGPT have recorded the lowest level of accuracy respectively. Overall, all the AI tools recorded 60% accuracy, 25% less accuracy, and 15% inaccuracy.

Keywords: Accuracy, AI tools, artificial intelligence, human translation, Iraqi dialect, literary texts

إمكانيات وحدود الذكاء الاصطناعي التوليدي في ترجمة اللهجة العراقية في الأدب: تقييم دقة ترجمة رواية "النوم في حقل الكرز" لأزهر جرجيس
م.د. رقية صبيح خلف الطائي
قسم الترجمة / كلية الآداب / جامعة البصرة
Ruqaya.sabeeh@uobasrah.edu.iq

المخلص

تناولت الدراسات الأكاديمية الحديثة استخدام أدوات الذكاء الاصطناعي في ترجمة النصوص الأدبية وأظهرت هذه الدراسات عموماً قدرة أدوات الترجمة المدعومة بالذكاء الاصطناعي على إنتاج ترجمات دقيقة للمفردات الشائعة. إلا أن أدائها قد يتباين عند التعامل مع التعبيرات الصعبة المرتبطة باللهجات والاختلافات الثقافية والمعاني السياقية. وفي الترجمة الأدبية من العربية، يُمثل الحفاظ على مستوى عالٍ من الدقة في ترجمة التعبيرات الثقافية تحدياً لأدوات الذكاء الاصطناعي. يهدف هذا البحث إلى تقييم مستوى دقة أربع ترجمات مدعومة بالذكاء الاصطناعي وهي شات جي بي تي (ChatGPT)، مايكروسوفت كوبيلوت (Microsoft Copilot)، ديب سيك (DeepSeek)، جيميني (Gemini) وترجمة بشرية واحدة لمفردات اللهجة العراقية في رواية "النوم في حقل الكرز" للكاتب العراقي أزهر جرجيس، والتي ترجمها جوناثان رايت إلى الإنجليزية. في هذه الدراسة، لن يقتصر البحث على تقييم دقة الترجمة فحسب، بل سيُقيم أيضاً أوجه التشابه والاختلاف

بين الترجمة البشرية وترجمات الذكاء الاصطناعي لل لهجة العراقية لتحديد أيهما أكثر موثوقية في فهمها. وقد اعتمد الباحث نظرية نوابان لدقة الترجمة لتحليل البيانات من حيث الدقة، وقلة الدقة، وعدم الدقة. وكشف التحليل أن مستوى الدقة كان متساوياً في الترجمة البشرية وديب سيك وجيمناي بنسبة ٦٧% ونسبة ١٦% عدم دقة في الترجمة، بينما سجل برنامجاً كوبايلوت وجات جي بي تي أدنى مستوى من الدقة. وبشكل عام، سجلت جميع أدوات الذكاء الاصطناعي الأربعة دقة بنسبة ٦٠%، وقلة دقة بنسبة ٢٥%، وعدم دقة بنسبة ١٥%.

الكلمات المفتاحية: الدقة، أدوات الذكاء الاصطناعي، الذكاء الاصطناعي، الترجمة البشرية، اللهجة العراقية، النصوص الأدبية

1. Introduction

The rapid development of artificial intelligence (AI) has altered the field of translation studies, presenting new translation tools, such as generative AI-driven language tools. Such technologies have become progressively indispensable in managing the increasing demand for quick and user-friendly translation across various languages and dialects, including the Arabic and English languages, since AI systems, according to Mohamed et al. (2024), use a variety of components, including data, neural networks, and linguistic expertise to effectively comprehend the intricacies inherent in many languages. In this regard, Khurana et al. (2022) elaborated that natural language processing (NLP) has been improved with the help of deep learning, particularly the use of neural networks with attention mechanisms. Its important first steps include preprocessing operations, such as tokenisation and stemming. NLP provides the backbone for language translation in the ever-evolving field of AI.

Hence, AI translation tools provide substantial advantages, such as quickness, consistency, and the capability to deal with huge amounts of texts, rendering them valuable for initial drafts and comparative analysis. Translation accuracy, for Mohamed et al. (2024), has attained new heights by reason of enhanced competences in understanding context and idiomatic expressions. Nevertheless, the translation of literary texts demands more than linguistic accuracy; it necessitates cultural awareness and a deep understanding of implicit meanings, predominantly those specific to culture. Professional human translators excel in identifying these culturally-specific expressions that might not have direct counterparts in the English language. Mohamed et al. (2024) confirmed that the combination of human creativity with AI output opens a never-ending vista of translation potential, contributing to the development of a global society capable of overcoming linguistic barriers. However, in literary translation contexts, rendering Arabic expressions, particularly those deeply rooted in Iraqi Arab culture, idioms, and social nuances, remains a complicated task.

In the current paper, the main research question is: to what extent do all AI tools, along with human translation, accurately understand the Iraqi-specific expressions in the novel? Further, how

does the accuracy of Wright's translation compare with that of the AI versions? Thus, the researcher will examine the human translation of Jonathan Wright¹ and the translations of the free versions of four AI tools, including ChatGPT, Copilot, DeepSeek, and Gemini, of Azher Jirjees'² Iraqi novel *At Rest in the Cherry Orchard*. When converting literary texts, especially this text which is rich in culture-specific Iraqi Arabic expressions, relying on a single method, whether human or AI, might limit the depth and accuracy of an interpretation of these expressions because they are related to the dialect of Iraq which is culturally rooted. Hence, having multiple AI translation options permits comparison between many linguistic choices and levels of cultural understanding. This mixture assists with highlighting nuances that might otherwise be disregarded, particularly when dealing with culturally-rooted connotations that do not have one-to-one equivalents in the English language. For this reason, when reviewing several AI outputs alongside human translation, literary translators of Iraqi literature can better assess how efficiently each version communicates the original Iraqi emotional and cultural context. While human translators bring cultural intuition and interpretive skills, AI systems offer speed, consistency, and alternative perspectives. Accordingly, a combination of these approaches ultimately leads to better-versed decisions and deeper, more accurate interpretations of Iraqi literary texts into the English language.

2. Theoretical background

2.1 Related Works

There have been a number of studies that have examined the effectiveness of AI tools in translating Arabic dialects, such as Emirati, Jordanian, Egyptian, etc. One of these studies was conducted by Al-Batineh et al. (2025) who, in their paper entitled *Potential Use of ChatGPT-4 for Translating the Emirati Dialect into English*, specifically focused on using ChatGPT to translate the Emirati dialect by analysing the translation of five texts, including 39 Emirati lexical terms into English. The study revealed that ChatGPT successfully understood the linguistic, cultural, and semantic complexities involved. This study differs from the current study in that it focuses on the Iraqi dialect and compares four AI translations with a human translation to check the validity of two machine translations.

In another research paper, *Metaphor Interpretation in Jordanian Arabic, Emirati Arabic and Classical Arabic: AI vs. Humans*, Zibin et al. (2025) examined how Jordanians and Emiratis, and four AI tools (ChatGPT-4, ChatGPT-3.5, Google Gemini, and Ask PDF), can recognise metaphors in

¹ Jonathan Wright is a prominent translator of Arabic literature and Iraqi literature. He is 'a British journalist and literary translator, known for bringing many works of Arab fiction to new audiences via translation for the past fifteen years. His recent works, however, seem more connected to Iraq' (Abuo Rashid, 2023).

² Azher Jirjees is an Iraqi novelist and writer. He has written many literary works, novel and short stories in Arabic and some of them have been translated into English including *At Rest in the Cherry Orchard*, *The Sweetmaker*, *The Stone Happiness*, *The Crow's Revolution*, and others.

classical Arabic and everyday Jordanian and Emirati dialects. The Mann–Whitney U model test was adopted to assess disparities in accuracy and interpretation between the AI tools and human participants in Jordan and the Emirates. The point of difference between this study and the present one lies in the sole emphasis of the latter on the Iraqi dialect, regardless of whether the expressions used are metaphorical or not and examines four AI tools to assess translation of the Iraqi dialect in addition to human translation.

In a paper, entitled, *Evaluating the Translation of Egyptian Colloquial Arabic Body Idioms into English: A Comparative Analysis of Four AI Tools*, Sadiq (2025) considered the efficiency of ChatGPT, Claude, Gemini, and DeepSeek in transforming Egyptian colloquial Arabic idioms into English based on the accuracy of meaning, naturalness, and cultural suitability. In contrast with the main objective of this study, it advocates a hybrid approach by joining AI competencies with the human touch. If we compare Sadiq's study with the current research, the latter deals with the translation of Iraqi expressions using four AI tools (ChatGPT, Copilot, DeepSeek, and Gemini) to assess the translation of an Iraqi literary novel encompassing Iraqi expressions into English.

2.2 AI Translation Tools

AI is a machine that makes use of technology to study and disentangle complications, being trained on huge volumes of information and able to distinguish patterns. At this time, AI tools and language models can be utilised together to recognise numerous inputs included in a variety of texts, images, audio recordings, videos, and PDFs (Imran & Almusharraf, 2024). AI uses large language models (LLMs) in the domain of Natural Language Processing (NLP) (Stefania, 2025). The advent of AI conversational interfaces, such as chatbots, focuses on the 'ability' of AI to accurately translate a wide range of (specialised) texts, and these chatbots, powered by LLMs trained on extensive data sets from books and the Internet, are designed to generate coherent and contextually relevant text by understanding and predicting language patterns (Stefania, 2025). The LLMs are deep learning-based AI models that analyse natural language to generate text, summarise information, answer questions, and translate between languages. These capabilities make LLMs ideal for powering AI chatbots, enabling them to interact conversationally, understand questions, and provide relevant answers (Tepe & Emekli, 2024a). Thus, all AI tools learn from massive amounts of data, including online texts, conversations, and images, keeping them up to date in terms of content. Furthermore, as highlighted by Tepe and Emekli, AI language models use a form of machine learning known as deep learning to process and predict the probability of each word or token in a sentence given the words that precede it. The models are trained on diverse datasets, including books, articles, and websites, to develop a comprehensive understanding of language patterns, grammar, syntax, and even context. The result is a powerful tool capable of a wide range of text-based tasks, from answering questions to translating

languages and assisting in specialised fields where accuracy and sensitivity to context are crucial (Tepe & Emekli, 2024b).

In the current paper, the researcher has selected four AI tools (ChatGPT, Gemini, DeepSeek, and Microsoft Copilot). Each tool will be described in terms of their distinguishing features and similarities and differences. Firstly, ChatGPT is an innovative AI model created by OpenAI, which represents a substantial progression in AI-assisted translation. As a sophisticated language model, ChatGPT has proven capable of delivering translations that are coherent, contextually relevant, and stylistically appropriate (Jiao et al., 2023). ChatGPT is built upon ChatGPT models and optimised with Reinforcement Learning from Human Feedback. Due to its surprising ability in natural language understanding and natural language generation tasks, ChatGPT has attracted more and more attention across millions of users (Brown et al., 2020).

Google Gemini's most prominent feature is its capability to recognise and deal with diverse data styles like texts, images, audios, PDFs, and videos as well. This AI tool features multiple LLMs and NLP technologies (Farrokhnia et al., 2023). Gemini encompasses different LLM sizes and incorporates three forms: Gemini Nano, Gemini Pro, and Gemini Ultra. These versions are thought to be consistent with users' requirements and demands. Nano is designed for 'on-device' efficiency and more accessible use on smartphones. Ultra fully uses Google's AI capabilities. The Pro version is balanced in terms of use and AI capabilities as it is a merged version of the other two (Team et al., 2025). In differentiating Google Gemini from ChatGPT, Gemini is integrally intended for multimodal operation, being capable of managing text, video, audio, code, and images.

DeepSeek was created by a Chinese AI research company which is renowned for evolving powerful LLM. Their flagship models comprise DeepSeek-V3, which is a general-purpose LLM with 671B parameters, and DeepSeek-R1, which is a reasoning-focused model that shows its thinking process. DeepSeek-V3 and then DeepSeek-R1 quickly gained global popularity after its release at the beginning of 2025 due to its claimed performance merits despite being an open-source model trained with low resources. DeepSeek R1; however, is a manifestation of a gradual improvement over numerous early models with the scaling law as its foundation (Guo et al., 2025). In contrast, ChatGPT and Gemini are closed-source AI models. By leveraging innovative training techniques, DeepSeek has disrupted the generative AI industry, becoming the benchmark for cost efficiency and openness for emerging startups (Thiebaut, 2026).

3. Methodology of the Study

The following translation prompt was given to the AI tools to translate the selected examples: *Translate the following text from Arabic into English and pay particular attention to Iraqi-specific*

terms. Four AI tools, including ChatGPT, Copilot, DeepSeek, and Gemini, and the human translation by Jonathan Wright, have been evaluated. The selected translation tools have been specifically adopted because they are widely used to translate words, sentences, texts, and even files, while the human translation was the only available translation of the novel. The examples extracted from the novel are not individual words or phrases of Iraqi dialect, but full sentences to give the AI tools the chance to understand the context and consequently allow for an evaluation of the accuracy of the translation output, comparing them to the human translation. The translation tools have been utilised between May 2025 and January 2026. Twelve examples from the Iraqi Arabic novel by Azher Jirjees, *At Rest in the Cherry Orchard*, have been extracted and categorised into three main themes: terms related to Iraqi food, terms related to Iraqi social expressions, and terms related to Iraqi humour and irony, with each of them covering four examples. Each theme encompasses four examples. In each example, the AI translations have been displayed (ChatGPT, Copilot DeepSeek, and Gemini) in addition to the human translation by Jonathan Wright in a separate table. These translations have been then analysed using Nababan's Translation Accuracy Theory and categorised as accurate, less accurate, or inaccurate.

This study implements a mixed-methods approach mingling qualitative analysis of the selected text with quantitative measures of numbers of accurate and inaccurate translations of each AI tool and the human translation. Nababan's Translation Assessment Model (2012) includes three pillars: accuracy, acceptability, and readability. However, in this study, the researcher has focused on the accuracy dimension. The instruments of accuracy in Nababan's theory encompass three main scales, namely accurate, less accurate, and inaccurate. Firstly, when the meaning of the source language's (SL) words, phrases, clauses, and sentences is accurately transferred to the target language (TL), and there is no omission, the translation is considered accurate. In other words, there is absolutely no distortion of meaning. With the less accurate scale, there are minor distortions of meaning in the TL that hinder the integrity of the meaning delivered by the SL. Thirdly, when the meanings of SL words, phrases, clauses, and sentences are not properly expressed into the TL, the translation is regarded as inaccurate (Nababan et al., 2012).

The 12 examples referenced above will be presented and analysed separately according to the aforementioned model. The main steps of analysing the three themes are as follows: identifying a selection of terms related to the Iraqi food, terms related to Iraqi social expressions and terms related to Iraqi humour and irony; tabulating the four translated versions of each of the excerpts selected for each theme; and then specifying the level of accuracy of each one.

4. Data Analysis

4.1. Terms Related to the Iraqi Food

Example 1

ST	نسيت اخراج كيس الكليجة من الحقيبة.	Category
ChatGPT	I forgot to take the bag of Kleija out of the bag.	Less accurate
Copilot	I forgot to take the bag of kleicha out of the bag.	Less accurate
DeepSeek	I forgot to take the bag of kliyacha out of my suitcase.	Less accurate
Gemini	I forgot to take the bag of Kleicha out of the suitcase.	Less accurate
Jonathan's	I had forgotten to take the bag of kleicha pastries out of my suitcase.	Accurate

Table 1: Iraqi Food 1

The Iraqi specific lexical item ‘Kleicha - الكليجة’ refers to a popular pastry baked during Eids (a Muslim religious holiday), Christmas, and other Iraqi cultural occasions. It is overstuffed with date paste, coconut, walnuts, or Iraqi Halqoom. *Kleicha* is most accurately rendered by Wright as ‘Kleicha pastries’ because of the addition of ‘pastries,’ since this addition helps the target reader to realise that *Kleicha* is a kind of pastry. All of the AI translation tools utilised can be considered as less accurate because they all chose to transliterate it as *Kleicha*, with some placing it in italics, while others presented it in bold and italics. In comparing the AI translations with Wright’s, it can be seen that they are less accurate because Wright specifically identified what *Kleicha* was to make it clear for the reader, while ChatGPT transliterated it only and left it to the reader to discover the exact meaning.

Example 2

ST	طلبت سندويش فلافل بالعنبة.	Category
ChatGPT	I ordered a falafel sandwich with 'Anaba (grape).	Inaccurate
Copilot	I ordered a falafel sandwich with pickled mango sauce.	Less accurate
DeepSeek	I ordered a falafel sandwich with amba .	Accurate
Gemini	I ordered a falafel sandwich with Amba .	Accurate
Jonathan's	I ordered a felafel sandwich with mango chutney	Less accurate

Table 2: Iraqi Food 2

The Iraqi lexical item ‘amba - عنبّة’ in the Iraqi context means a kind of spicy sauce consisting of ground small seeds, fenugreek, turmeric, lemon, and vinegar. It is used in falafel sandwiches. With Google, it was translated as ‘grapes’ and this is inaccurate because it does not only fail to convey the meaning of the original ‘amba,’ but also it does not make sense to the reader to have falafel with grapes. As for ChatGPT, the word was transliterated and the meaning provided between brackets as ‘grape.’ The meaning was lost because ‘grape’ is not an accurate rendition, nor is it related to sandwich sauces. Regarding DeepSeek and Gemini, their translations were considered as being less accurate because on the one hand, the word ‘amba’ might be unclear for TL receivers in English since ‘amba’ is an Iraqi-specific item related to food. On the other hand, since it was written in bold and the literal translation may not provide an inaccurate rendition in English, the aforementioned AI tool decided to transliterate it, leaving it to the target audience to check the exact meaning of ‘amba.’ Further, Copilot rendered it as ‘sauce,’ making some sense, since people always use sauce with falafel, and mango sauce is one of the most popular sauces used, but it was not the intended meaning in the original text. As far as Wright’s translation is concerned, it was also considered less accurate due to the reason that mango chutney can be eaten with sandwiches in Iraq, but *amba* has a totally different meaning. In my opinion, ‘amba sauce’ would have been the most accurate translation, with the addition of sauce indicating that amba references sauce, not anything else.

Example 3

ST	كانت توقظني أيام الجمع مبكرا كي أحضر <u>الصمّون</u> من الفرن القريب من بيتنا	Category
ChatGPT	She used to wake me up early on Fridays so that I would go and buy samoon from the bakery near our house. In	Accurate
Copilot	She used to wake me up on Fridays so I could bring samoon from the bakery near our house.	Accurate
DeepSeek	Every Friday, she would wake me up early so I could get fresh samoon from the bakery near our house. In	Accurate
Gemini	She used to wake me up early on Fridays so I could fetch Samoon from the bakery near our house. In	Accurate
Jonathan’s	She would wake me up early on Fridays to fetch loaves of bread from the bakery near our house.	Less accurate

Table3: Iraqi Food 3

If one looks at the four AI translations, they translated the Iraqi lexical item ‘samoon’ as ‘samoon,’ except for DeepSeek which referenced it as ‘fresh samoon.’ The latter transliterating this lexical term leaving it to the target audience to guess and search its meaning, since this term does not exist in English. Moreover, ‘fresh samoon’ can be considered accurate compared with *samoon* because the addition of ‘fresh’ gives some definite indication that it is a kind of bread, and it may be somehow that ‘samoon’ is related to bakery and indicates that it is a kind of bread in Iraq called *samoon*. Conversely, Google rendered it simply as ‘bread.’ Bread is a very general description of this type of food that every country eats. Iraq, like all other countries, has different types of bread, and *samoon* is one of them. On the other hand, if one looks at the human translation by Wright, he rendered it as ‘loaves of bread,’ which partially preserved the meaning, but he did not specify what kind of bread it was, and there are different types of bread in Iraq. In addition, loaves of bread are different from the names of breads in Iraq, with *samoon* being one of them. In the researcher’s opinion, the most accurate translation would have been ‘samoon bread’ because the addition of ‘bread’ implies that it is a kind of Iraqi bread called *samoon*.

Example 4

ST	واحد لبن من فضلك.	Category
ChatGPT	One laban, please.	Less accurate
Copilot	One yogurt drink, please.	Accurate
DeepSeek	One [bottle of] laban, please.	Less accurate
Gemini	One Laban , please.	Less accurate
Jonathan’s	I’ll have a yoghurt too, please.	Inaccurate

Table 4: Iraqi Food 4

Regarding the fourth example, it can be seen that all of the translations are slightly different. ‘Laban’ is fermented milk and a yogurt-like drink. It is softer than yogurt, and that is why it is considered a drink in Iraq. Subsequently, the transliteration of the term as ‘laban’ can be considered slightly accurate because it is a culture-specific term in Iraq. On the other hand, in other cultures like Egypt, ‘laban’ refers to milk. Copilot can be regarded as accurate because it used the phrase ‘yogurt drink,’ indicating that it is a yoghurt that is not eaten but drunk. The human translation can be considered inaccurate because Wright rendered it as yogurt, and it should be drunk, not eaten. On the other hand, the rest of the AI platforms rendered it as ‘laban,’ differentiating the word by putting it in either capital letters or bold. Also, the addition of ‘bottle’ in DeepSeek can be regarded as a good

addition because it clarified the denotation of the word ‘laban,’ even if it was not aware of its exact meaning, because it indicates that it is something to be drunk, not eaten. Thus, the DeepSeek translation can be regarded as the most accurate among those which used the lexical item ‘laban.’

4.1. Terms Related to Iraqi Social Expressions

Example 1

ST	فيهر رأسه وقال كلمته الشهيرة: (عفية).	Category
ChatGPT	Then he nods his head and says his famous word: 'Afiya' (meaning 'Well done' or 'Good job').	Accurate
Copilot	He nodded his head and said his famous word: ‘Good job.’	Accurate
DeepSeek	He shook his head and said his famous line: "'Afya!"	Inaccurate
Gemini	So he shook his head and said his famous word: (Afiya / Well done).	Accurate
Jonathan's	When he nodded his head and said ‘Well done’, his favourite expression. (Jonathan, 34)	Accurate

Table5: Iraqi Social Expression 1

The lexical term ‘afya – عفية’ is used to praise someone for their efforts when they are doing something good, meaning ‘well done,’ ‘good job,’ ‘you did great,’ etc. If one looks at the AI translations alongside Wright’s, it can be perceived that they all produced an accurate translation, except for DeepSeek, which rendered it as ‘Afia,’ and Google, which translated it as ‘may you be well.’ The Google translation was also inaccurate because it gave ‘may you be well’ and the translation of ‘afya,’ which is the transliteration of the term, and this strategy is pointless for the target reader. The ChatGPT translation conveyed however an accurate rendition of the meaning in English as ‘good job,’ whereas Wright’s translation was ‘well done.’ Thus, ChatGPT’s translation can be categorised as an accurate translation.

Example 2

ST	تفضل يا سيدي، على راسي.	Category
ChatGPT	Please, sir, it's an honor.	Less accurate
Copilot	Please, sir, it's my honor.	Less accurate
DeepSeek	Of course, sir—it's my pleasure / at your service.	Accurate
Gemini	Go ahead, sir; [you are] on my head .	Inaccurate
Jonathan's	There you are, sir, at your service.	Accurate

Table 6: Iraqi Social Expression 2

The expression ‘ala rasi- على راسي’ is used to convey humbleness and respect to people when providing a service or offering something to someone. When looking at all of the translations, it can be seen that they are different. For instance, Gemini produced a totally literal translation which was inaccurate and ambiguous and had no relation to the context, being ‘on my head.’ On the other hand, in considering the translations of ChatGPT and Copilot, they rendered it as ‘it’s an honor,’ denoting that it shows respect, but the meaning of the original phrase does not mean ‘it’s an honor.’ Conversely, the translations of DeepSeek and Wright can be regarded as accurate translations of the original, being ‘at your service’ and ‘it’s my pleasure’ respectively.

Example 3

ST	هتف به المثلث: (هويتك.. <u>خلصني</u>).	Category
ChatGPT	The masked man shouted at him: “Your ID—hand it over.	Accurate
Copilot	The masked man shouted at him: ‘Your ID... give it to me.’	Accurate
DeepSeek	Your ID... hurry up and give it to me!	Accurate
Gemini	Your ID.. hurry up / get it over with).	Accurate
Jonathan’s	“Your ID card, please. Let’s get this over with,” said the man.	Accurate

Table 7: Iraqi Social Expression 3

The Iraqi lexical item ‘خلصني’ largely means ‘get it over,’ ‘get rid of something,’ or ‘finish it,’ and in this context it is said with anger. If one looks at all of the translations of the same Iraqi-specific term, it is evident that the human and AI translations can be considered accurate except for that of Google which was less accurate. The translation by Google as ‘save me’ can be interpreted by target text receivers as ‘get it over,’ or literally ‘save me.’ That is why, it can be regarded as less accurate, because the literal meaning of ‘save me’ can mislead the reader. On the other hand, the rest of the translations maintained the meaning of the original with different styles of expression, such as ‘hand it over,’ ‘give it to me,’ ‘get it over,’ and ‘let’s get this over.’

Example 4

ST	فقال: (على عيني.) وأضاف لي ثقبين بدل الثقب الواحد، تحسباً لفقد أرتال أخرى.	Category
ChatGPT	He said, “Consider it done.” And he added two holes instead of one, just in case more pounds were lost later.	Accurate
Copilot	He said: ‘With pleasure,’ and added two holes instead of one, anticipating the loss of more pounds.	Accurate
DeepSeek	So he said, "My pleasure," and added two holes for me instead of one, in anticipation of losing more weight.	Accurate
Gemini	So he said: (On my eye / With pleasure.) And he added two holes [to the belt] instead of just one, in anticipation of losing more pounds.	Accurate
Jonathan’s	“Sure,” he said. In case I lost more weight he punched two new holes in the belt, rather than just one.	Accurate

Table 8: Iraqi Social Expression 4

In Iraq, the expression ‘على عيني’ is meant to refer to phrases like ‘of course,’ ‘sure,’ ‘my pleasure,’ or ‘consider it done.’ When a person asks someone else to do something or has a request, the reply is usually ‘على عيني.’ In Iraqi culture, the word *eye* is frequently used, and it represents caring for people and being honoured to do something for them. Thus, ‘على عيني’ literally means ‘on my eye,’ and this denotes that you will treat the request with great responsiveness and respect, and you are glad to do it for them. Thus, all of the translations can be regarded as accurate renditions of the original in terms of conveying the meaning.

4.2. Terms Related to Iraqi Humour and Irony

Example 1

ST	وصل بابا عفلق استعدوا للخرط	Category
ChatGPT	Papa Aflaq has arrived, get ready for the nonsense.	Accurate
Copilot	Papa Aflaq has arrived — get ready for nonsense.	Accurate
DeepSeek	Baba 'Aflag has arrived, get ready for the [political] uproar.	Accurate
Gemini	Papa Aflaq has arrived; get ready for the nonsense [bullsh*t].	Accurate
Jonathan’s	Papa Aflaq has arrived, get ready for the bullshit”	Accurate

Table 9: Iraqi Humour and Irony 1

In Iraqi culture, ‘lilkharit – للخرط’ means talk that is overstated, telling unrealistic stories, or lying. People in Iraq also use it in a humorous way when someone exaggerates and pretends to know something they do not. In the political context, and in a comic style, Iraqis describe political speeches in this way if they do not believe or agree with the person, view their speeches as empty and comprising false promises, or when there is no real action. Therefore, its literal meaning can be ‘nonsense,’ ‘bullshit,’ etc. If we look at the translations of this lexical item, it is obvious that ChatGPT, Copilot and Gemini have accurately rendered the meaning as ‘nonsense.’ Similarly, Wright’s translation can be regarded as an accurate rendition of the original as ‘bullshit.’ Also, DeepSeek provided a translation that can be considered accurate because of the addition of ‘political’ to ‘uproar,’ referring to the fact that Papa Aflaq is an Iraqi politician.

Example 2

ST	تستهزئ جنابك!؟	Category
ChatGPT	Are you mocking me, sir?!	Less accurate
Copilot	Are you mocking me, sir?	Less accurate
DeepSeek	Is Your Excellency making fun of me?	Accurate
Gemini	Are you mocking [me], your excellency!?	Accurate
Jonathan’s	Is your highness making fun of me?!	Accurate

Table 10: Iraqi Humour and Irony 2

In the Iraqi context, the lexical item ‘jenabek- جنابك’ can be used in two different ways. The first is when it is used formally to express respect to somebody, and the second when it is used to express the opposite, especially rejection, disrespect, insult, etc. In this latter context, it is not a formal or polite method of expression. Consequently, the ChatGPT and Copilot translations can be perceived as less accurate because they did not convey a clear rejection of making fun of somebody by using ‘sir,’ since ‘sir’ represented respect to the receiver in this context. Conversely, Wright’s translation, in addition to that of Gemini and DeepSeek, added ‘your highness’ and ‘your excellency,’ indicating that the person was not pleased by this fun. In this case, they can be regarded as accurate renditions of the meaning of the original. Similarly, despite Google deleting the translation of ‘جنابك,’ it conveyed direct rejection by saying: ‘Are you kidding me?’

Example 3

ST	لا تزعل مني انا امزح معك يا صديقي.	Category
ChatGPT	Don't be upset with me, I'm just joking with you, my friend.	Accurate
Copilot	Don't be upset with me, I'm just joking with you, my friend.	Accurate
DeepSeek	Don't be upset with me, I'm just joking with you, my friend.	Accurate
Gemini	Don't be upset with me; I am joking with you, my friend.	Accurate
Jonathan's	Don't get upset, I am just kidding.	Accurate

Table 11: Iraqi Humour and Irony 3

The Iraqi Arabic expression ‘لا تزعل مني’ is popularly used in the Iraqi dialect, and its accurate meaning is ‘do not be upset with me.’ It can be said after doing something that might annoy others or when making a joke, or teasing another, and it can also be considered as a soft apology. In examining the translations of the lexical item ‘la tiz'al mini - لا تزعل مني,’ it can be seen that they are accurate because they all translated the meaning as ‘Don't be upset with me’ in addition to the human translation of ‘Don't get upset.’

Example 4

ST	قال حمزة الأملط وهو يوصلني عند باب المطار.	Category
ChatGPT	Hamza al-Amalt said as he dropped me off at the airport gate...	Inaccurate
Copilot	Hamza the bald said this while dropping me off at the airport gate.	Inaccurate
DeepSeek	Hamza Al-Amlat said as he dropped me off at the airport terminal door:	Inaccurate
Gemini	Hamza Al-Amlat said as he was dropping me off at the airport gate.	Inaccurate
Jonathan's	Hamza al-Amalt said as he dropped me off at the airport gate. (Jonathan, 58)	Inaccurate

Table 12 : Iraqi Humour and Irony 4

In the Iraqi context, the lexical item ‘الاملط’ means someone who is hairless in the face; thus, they have no beard or, particularly, no moustache. It can also refer to a person without body hair, but in this context, it references a beardless person or one without a moustache. The expression ‘Hamza Al-

Amlat - حمزة الأملط' was inaccurately rendered by Wright, ChatGPT, DeepSeek, Gemini, and Google. The mistake in translation is related to the transliteration of 'Al-Amlat.' All of the translations rendered it as a surname when, in fact, it is an adjective used by Iraqis to describe a man with no beard. It is not the surname of a person; there is no Al-Amlat tribe in Iraq. On the other hand, if we look at Copilot's translation, it can be seen that it rendered the item as 'blad,' referring to a person who has no hair on his head, and this is not what was intended in the original. This is due to the fact that when an Iraqi wants to refer to a person who has a hairless head, they call him 'الأصلع,' not 'الأملط.'

5. Data Analysis and Discussion

The researcher has conducted a qualitative study assessing the translation quality of four AI platforms, including ChatGPT, Microsoft Copilot, DeepSeek, and Gemini, and one human translation by Jonathan Wright of Azher Jirjees' novel *At Rest in the Cherry Orchard*. The translations of several Iraqi cultural terms have been assessed and discussed. These five translations (from four AI tools and one human translation) have been assessed in terms of being accurate, less accurate, or inaccurate. The percentages of accuracy of the AI translation tools and the human translation are calculated in the following figures separately in **Figure 1**, **Figure 2**, and **Figure 3**.

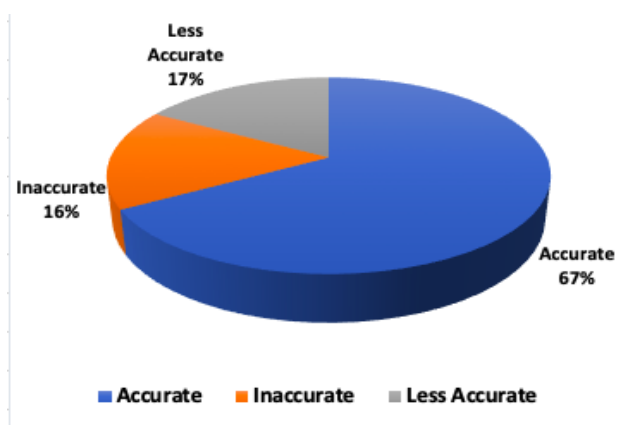


Figure 1: DeepSeek Translation

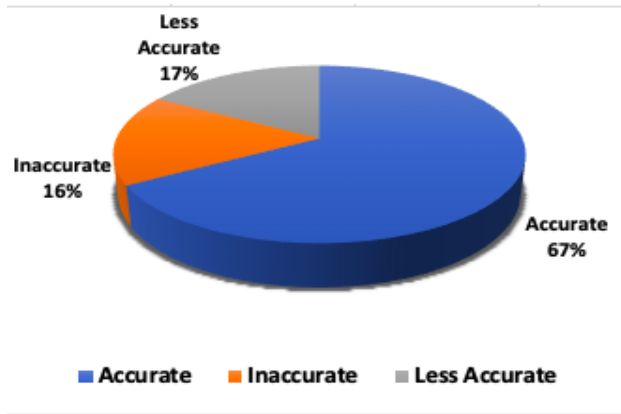


Figure 2: Gemini Translation

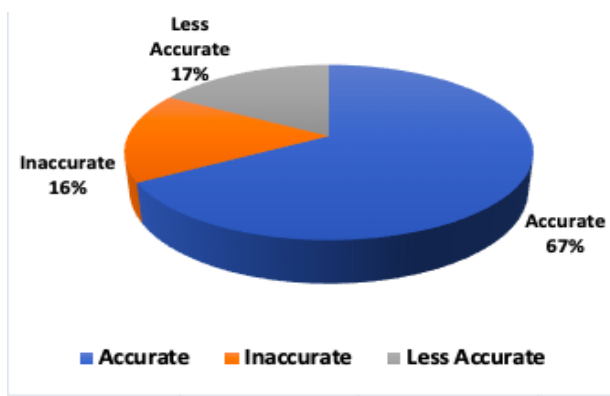


Figure 3: Jonathan's Translation

In the above figures, since the percentages are similar, the researcher decides to put them together for analysis, and then examines the rest afterwards. In looking at the three figures, it is obvious that the percentages for DeepSeek, Gemini and Wright's translation are equal, rendering all 12 examples with the highest rate of accurate translation at 67%, with 17% less accurate, and 16% inaccurate. The similarity of percentages between Wright's translation and the AI translations by DeepSeek and Gemini indicate that the level of AI accuracy of translation is high, and AI tools can be used as assistant tools for human translators when dealing with Iraqi-specific cultural items with the aim of understanding their meaning in Arabic and consequently translating them into English. One question may arise: why is the human translation not higher than that of the AI tools? This may be because, despite the human translator being a professional translator, the inaccurate scale is very low (16%). Wright was unsuccessful in conveying only two terms. This is due to these terms being confusing and too much rooted in Iraqi culture, and their interpretations being treated differently in other Arabic dialects (this includes terms such as 'laban' and 'al-amlat'). Regarding the AI tools, and

which are more accurate and reliable in translating the Iraqi tongue, it can clearly be perceived that Gemini and DeepSeek are more reliable than ChatGPT and Copilot (see **Figure 4** and **Figure 5**).

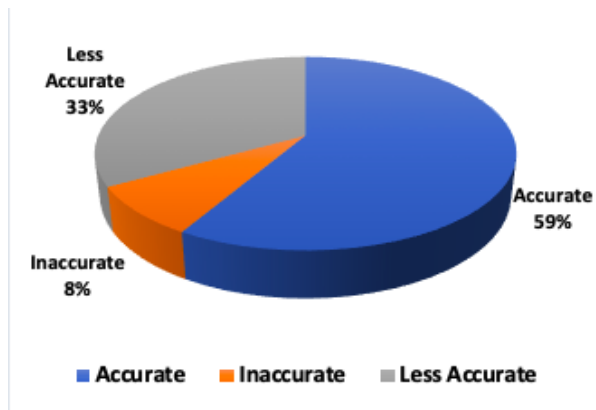


Figure 4: Copilot Translation

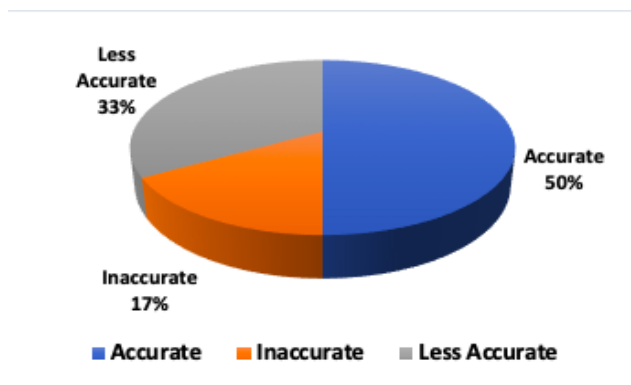


Figure 5: ChatGPT Translation

1.

Regarding Copilot and ChatGPT, Copilot's translation represents an accurate rendition of all the examples; it recorded the highest rate 59%, while ChatGPT has 50% accurate rendition and 50% less accurate and inaccurate. Copilot has recorded the lowest rate of all the AI tools, while the human translation has recorded inaccurate translations at 8% only. This means that Copilot is the third most reliable tool for translation after Gemini and DeepSeek. On the other hand, ChatGPT has recorded the highest rate of inaccurate translation, while the human translation at 17% is quite similar to Gemini (16%), DeepSeek (16%), and human translation (16%). For the less accurate percentage, both Copilot and ChatGPT have recorded the same rate (33%). If one compares this rate with DeepSeek, Gemini, and Wright's translations, it can be seen that the less accurate rates are lower than that of Copilot and ChatGPT at 17%. Again, this indicates that DeepSeek and Gemini, as AI translation tools, are performing at a high level.

After discussing all of the translations, the researcher has found that it is important to combine all of the translations together and calculate the total number of accurate, inaccurate and less accurate translations of the tools collectively (incorporating both the AI tools and human translation) to check the total rates of them as a whole (see **Figure 6**, **Figure 6** **Figure 8**, and **Figure 8**).

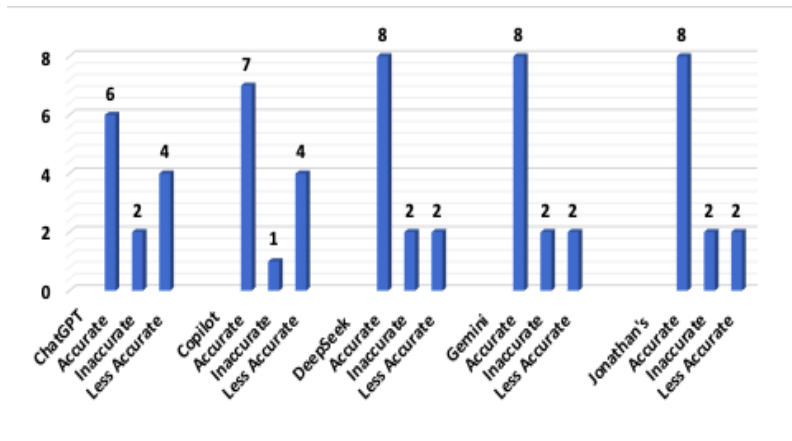


Figure 6: All Translations Together (AI and HT)

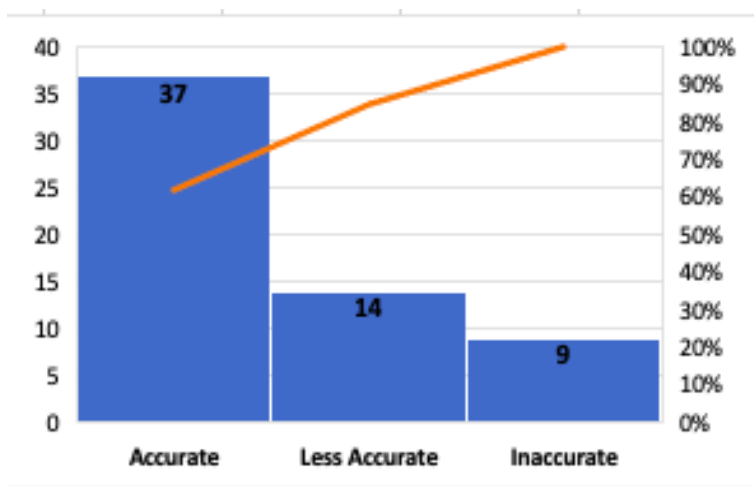


Figure 7: The Total Numbers (AI and HT)

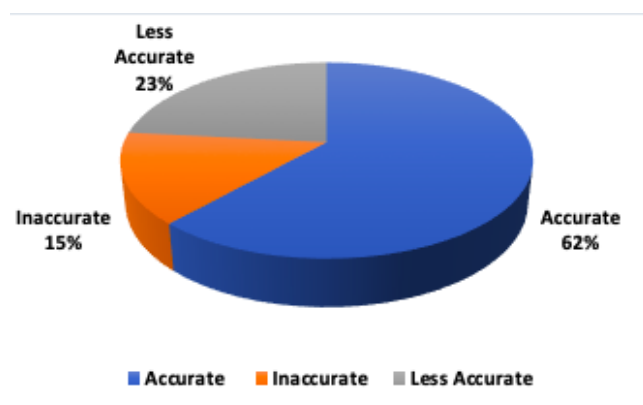


Figure 8: The Whole Percentage (AI and HT)

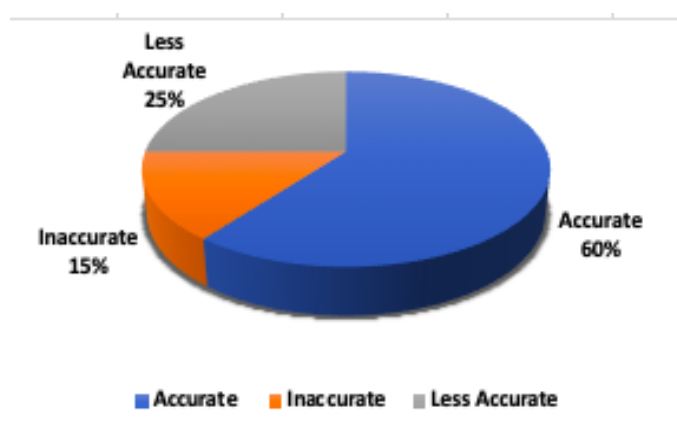


Figure 9: The Whole percentage (AI Tools)

The total number of accurate, inaccurate, and less accurate translations of the 12 examples is 60. In this research, as mentioned previously, there are 12 examples; each example comprises five translations. Therefore, 12 multiplied by 5 equals 60 translations. The total number of accurate renditions of the original Iraqi-specific terms by all translation tools and the human translation is 37, occupying 62%, while the inaccurate and less accurate numbers of translation have recorded 9 (15%) and 14 (23%) respectively. This means that the accurate percentage has occupied more than 62%, with the AI tools recording 29 accurate translations, occupying 60%, while the inaccurate and less accurate translations have accounted for 7 (15%) and 14 (25%) respectively, as illustrated in the above graphs. This means that 85% of them are considered accurate or less accurate, while only 15% are inaccurate. Therefore, this percentage indicates that these AI tools are demonstrating high levels of accuracy in understanding the meaning of Iraqi-specific cultural lexical items.

2.

6. Conclusion

Numerous academic studies investigated the performance of several AI translation tools in rendering texts from Arabic into English. In general, these tools demonstrated acceptable levels of accuracy in transferring the general meaning and Iraqi-specific terms encompassed in Iraqi Arabic literary texts. Nevertheless, paying careful attention to contextual interpretations of Iraqi literary texts is required, and AI systems may or may not fully capture the accurate meaning in English. Moreover, the accuracy of these tools varies. The main research question of this paper is: how accurately do all AI translations, including HT, capture the meaning of Iraqi cultural expressions? The second question was: Further, how much more accurate is Wright's translation than the AI versions? The main answer to this question is that all of the AI tools and the human translation have managed to accurately translate many of the Iraqi terms into English. DeepSeek, Gemini, and Wright have received equal percentages of around 67% accuracy in translation, meaning that these AI tools have reached high levels of accuracy in understanding the meaning of Iraqi expressions and can be utilised as assistant tools by professional human translators in dealing with Arabic expressions. Consequently, AI technologies are commonly being regarded as supportive tools that can assist translators rather than replacing the essential role of human expertise. Thus, AI is not a substitution for translators, but can be adopted as an assistive tool to support human translators and improve efficiency, ultimately leaving interpretative judgements to human expertise. Therefore, the use of AI tools alongside human translation is not only practical, but necessary. By comparing AI-generated translations with human interpretations, translators can better evaluate the strengths and limitations of each tool. This combined perspective will enhance the quality of literary translation and contribute to a deeper cross-cultural understanding between Arabic and English audiences.

3. References

- Al-Batineh, M.; Al Tenaijy, M.; Sharkas, H.; Alblooshi, S.; Alsenaani, F.; Alquodhi, A. (2025). Potential Use of ChatGPT-4 for Translating the Emirati Dialect into English. *Theory & Practice in Language Studies (TPLS)*, 15(9) 2793-2801, <https://doi.org/10.17507/tpls.1509.01>
- Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2023). "A SWOT Analysis of Generative AI: Implications for Educational Practice and Research". *Innovations in Education and Teaching International*, 3(61), 460-474. <https://doi.org/10.1080/14703297.2023.2195846>
- Guo, Daya; Yang, Dejian; Zhang, Haowei; Song, Junxiao; Zhang, Ruoyu; Xu, Runxin; Zhu, Qihao; Ma, Shirong; Wang, Peiyi; Bi, Xiao, et al. (2025). "Deepseek-r1: Incentivizing Reasoning Capability in llms via Reinforcement Learning".

<https://doi.org/10.1038/s41586-025-09422-z>

- Imran, M.; Almusharraf, N. (2024). "Google Gemini as a Next Generation AI Educational Tool: A Review of Emerging Educational Technology". *Smart Learn. Environ*, 11(22), 1-8. <https://doi.org/10.1186/s40561-024-00310-z>
- Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). "Is ChatGPT a Good Translator? Yes with GPT-4 as the Engine". <http://arxiv.org/abs/2301.08745>
- Jirjees, A. (2019). *النوم في حقل الكرز (al-nowm fi haql al-karaz)*. Beirut: Dar Al-Rafidain.
- Jirjees, A. (2024). *At Rest in the Cherry Orchard*. Jonathan Wright (Trans.). London: Banipal Books.
- Khurana, D; Koli, A.; Khatter, K.; Singh, S. (2022). "Natural Language Processing: State of the Art, Current Trends and Challenges". *Multimedia Tools Appl.*, 82, 3713-3744. 10.1007/s11042-022-13428-4
- Mohamed, Y. A.; Khanan, A.; Bashir, M.; Mohamed, A. H. H. M.; Adiel, M. A. E. and Elsadig, M. A. (2024). "The Impact of Artificial Intelligence on Language Translation: A Review". *IEEE Access*, 12, 25553-25579. 10.1109/ACCESS.2024.3366802
- Nababan, M., Nuraeni, A., & Sumardiono, S. (2011). "Pengembangan Model Penilaian Kualitas Terjemahan". *Kajian Linguistik dan Sastra*, 24(1), 39-57. [oai:generic.eprints.org:12122/core478](http://oai.generic.eprints.org:12122/core478)
- Sadiq, S. (2025). "Evaluating the Translation of Egyptian Colloquial Arabic Body Idioms into English: A Comparative Analysis of Four AI Tools". *Egyptian Journal of Linguistics and Translation*, 15(1), 1-36. 10.21608/ejlt.2025.378658.1109
- Stefania M, M. (2025). "The Impact of Artificial Intelligence (AI) on Translation: An Overview". *International Journal of Language Studies*, 19(2), p147. <https://hdl.handle.net/10446/300447>
- Team, G., Anil, R., Borgeaud, S., Wu, Y., Alayrac, J. B., Yu, J., et al. (2025). "Gemini: A Family of Highly Capable Multimodal Models". *ArXiv*. <https://arxiv.org/pdf/2312.11805>
- Tepe, M. and Emekli, E. (2024a). "Decoding Medical Jargon: The Use of AI Language Models (ChatGPT-4, BARD, Microsoft Copilot) in Radiology Reports". *Patient Education and Counseling*, 126. <https://doi.org/10.1016/j.pec.2024.108307>
- Tepe, M; Emekli, E. (2024b). "Assessing the Responses of Large Language Models (ChatGPT-4, Gemini, and Microsoft Copilot) to Frequently Asked Questions in Breast Imaging: A Study on Readability and Accuracy". *Cureus*, 16(5). [10.7759/cureus.59960](https://doi.org/10.7759/cureus.59960)
- Thiebaut, R. (2026). "Evaluating the sudden success of DeepSeek". *Sage Business Cases*. SAGE Publications, Ltd. <https://doi.org/10.4135/9798348833091>

Brown, Tom; Mann, Benjamin; Ryder, Nick, Subbiah, Melanie; et al. (2020). “Language Models are Few-Shot Learners”. *Advances in neural information processing systems*.

[//doi.org/10.48550/arXiv.2005.14165](https://doi.org/10.48550/arXiv.2005.14165)

Zibin, A.; Binhaidara, N.; Al-Shahwan, H.; Yousef, H. (2025). “Metaphor Interpretation in Jordanian Arabic, Emirati Arabic and Classical Arabic: Artificial Intelligence vs. Humans”. *Humanit Soc Sci Commun*, 12(942). <https://doi.org/10.1057/s41599-025-05282-0>