

AI Tools in Translation: Bridging Literalism and Cultural Fluency in English-Arabic Idioms – A Comparative Analysis of ChatGPT and DeepSeek

Shaimaa Mohamed Helal¹

¹ Faculty of Language Studies, Arab Open University, Saudi Arabia. ORCID ID: <https://orcid.org/0009-0003-1540-7188>.

Correspondence: Shaimaa Mohamed Helal, Faculty of Language Studies, Arab Open University, Saudi Arabia.

Received: October 20, 2025

Accepted: January 20, 2026

Online Published: June 11, 2026

doi:10.5430/wjel.v16n5p290

URL: <https://doi.org/10.5430/wjel.v16n5p290>

Abstract

Idioms are fixed phrases whose forms are usually unchangeable and whose meanings are always the same, metaphorical, indirect and inflexible. They are considered as cornerstone of linguistic and cultural nuance. Therefore, their translation poses formidable challenges for AI-driven machine translation. The aim of the current study is to evaluate the performance of two leading AI tools, ChatGPT and DeepSeek, in translating 20 English idioms into Arabic. The study employs two theoretical frameworks from translation studies—Nida's dynamic equivalence as well as Vinay and Darbelnet's translation strategies—alongside principles of neural machine translation (NMT). The analysis is based on three criteria. These are (1) accuracy, (2) cultural appropriateness, and (3) naturalness. Through this tripartite analysis, the findings of the study reveal distinct approaches of the AI tools under scrutiny. DeepSeek frequently employs modulation as well as culturally adapted alternatives. This aligns closely with Nida's emphasis on meaning preservation over literalism. In contrast, ChatGPT often defaults to literal translations of meaning. Thus, it risks eroding idiomatic essence. The study concludes that translators must critically evaluate outputs to ensure cultural fidelity. For example, a translator must critically evaluate DeepSeek's literal translation of the idiom "signing on the dotted line" as ("وقعوا على الخط المنقط") and modify it to be ("وقعوا على بياض") to suit Arabic readers. Furthermore, the study advocates for a collaborative human-AI workflow (i.e., AI generates baseline translations and humans refine them through cultural vetting as well as stylistic adjustments). The study recommends that future advancements in NMT systems must prioritize idiom-specific training datasets and culturally annotated corpora, in addition to integrating dynamic equivalence principles with adaptive strategies with the aim of enhancing cultural fluency. Thus, by bridging algorithmic precision with human intuition, AI can evolve into a culturally intelligent partner that empowers translators.

Keywords: AI translation, idioms, ChatGPT, DeepSeek, Arabic translation, dynamic equivalence, cultural adaptation, neural machine translation, human-AI collaboration

1. Introduction

Idioms are phrases whose meanings cannot be deduced from the literal definitions of the words they comprise. Thus, they are deeply embedded in cultural contexts. Furthermore, they are often non-translatable using direct linguistic equivalents (Baziotis et al., 2023). This cultural embeddedness, as argued by Haider et al., (2025), often renders them non-translatable by using direct linguistic equivalents. Additionally, their meaning is figurative and cannot be derived from the meaning of their constituents (Nunberg et al., 1994; Liu, 2017; Baziotis et al., 2023). Therefore, idioms present a significant challenge for cross-linguistic communication (Baziotis et al., 2023; Haider et al., 2025). In the current age of artificial intelligence, advanced neural machine translation (NMT) tools as well as large language models (LLMs), such as ChatGPT and DeepSeek, have become increasingly prevalent in both professional as well as casual translation contexts (Liu et al., 2023). AI has been applied in translating various text types (Guo & Amirdabbaghian, 2025). However, despite their remarkable progress in handling literal texts, idiomatic expressions remain a major open problem for these systems (Baziotis et al., 2023). The core challenge lies in the models' tendency to default to literal (word-for-word) translations of idioms. This leads to confusing, nonsensical, or semantically lossy outputs in the target language (Zaitova et al., 2025). The reason is that these systems often struggle to disambiguate between the literal and figurative meanings of a phrase without sufficient contextual understanding (Liu et al., 2023). Therefore, the translation of idioms necessitates a strategy that prioritizes dynamic equivalence, (Nida, 1964), —where the focus is on reproducing the same effect on the target reader as the original had on the source reader—over formal equivalence, which is a more literal approach. This often requires the substitution (or replacement) of the source-language idiom with a culturally appropriate analogue in the target language. This is a complex task. It demands deep cultural and pragmatic knowledge (Vinay & Darbelnet, 1995).

The current study aims to systematically explore the strengths as well as the limitations of contemporary AI models. These models are ChatGPT and DeepSeek. The two models will be examined in understanding, interpreting, and translating twenty idiomatic expressions from English into Arabic. The rationale behind selecting Arabic is that Arabic is a language replete with its own rich as well as complex set of idiomatic expressions. Thus, it is a potent case study for examining how AI navigates profound cultural and linguistic nuances.

2. Objectives of the Study

The main objective of the current study is to conduct a comparative analysis of Arabic idiom translations generated by ChatGPT and DeepSeek. The focus is on evaluating their (1) accuracy, (2) cultural appropriateness, and (3) naturalness through established translation frameworks. Furthermore, the study applies Nida's dynamic equivalence theory to assess how effectively each AI system preserves both meaning and cultural context. Additionally, the study examines the use of Vinay and Darbelnet's translation strategies in machine-generated outputs and evaluates the performance of Neural Machine Translation (NMT) systems in handling idiomatic expressions as well as culturally embedded phrases.

From a practical standpoint, the study aims to identify system strengths, such as ChatGPT's consistency in literal translations and DeepSeek's cultural adaptation. Also, the study documents common failure points. These appear particularly when translating culturally specific idioms without direct equivalents. Furthermore, the study proposes strategies for improving NMT systems (particularly in the areas of metaphor detection, contextual awareness, and register matching).

In terms of educational objectives, the study seeks to create a classified dataset of 20 examples. It categorizes them into perfect translations, culturally adapted versions, and problem cases (which require human intervention). Finally, the study aims to develop guidelines for educators teaching Arabic-English idiom translation as well as for translators utilizing AI tools.

3. Research Questions

To achieve the goals of the study, the present study attempts to answer the following questions:

1. How accurately can AI tools translate English idioms into Arabic while preserving their intended meaning?
2. To what extent do AI translations reflect culturally appropriate and idiomatically equivalent expressions in Arabic?
3. How natural and fluent are the Arabic translations produced by AI tools from the perspective of native speakers?
4. What are the comparative strengths and limitations of ChatGPT and DeepSeek in the idiomatic translation task?

4. Statement of Research Problem

The current study seeks to systematically compare the performance of ChatGPT and DeepSeek in handling English-Arabic idiom translation. It examines their respective strengths and weaknesses through the lenses of (1) dynamic equivalence, (2) translation strategies, and (3) natural language fluency. The research addresses critical gaps in existing literature. These are: (1) the lack of comparative studies on Arabic idiom translation in these platforms, (2) the need for strategy-based evaluation frameworks, and (3) the absence of clear guidelines for post-editing machine-translated idioms. Thus, the current study aims to contribute both practical improvements for neural machine translation (NMT) development and theoretical insights into the complexities of cross-cultural idiom translation.

5. Literature Review

5.1 ChatGPT and DeepSeek

Within the fast-paced development of artificial intelligence (AI), natural language processing (NLP) models have become instruments of significant change for written composition, communication, and translation. ChatGPT is an advanced AI language model developed by OpenAI and DeepSeek is a powerful AI-powered search and discovery tool by a Chinese AI innovator creating open-source, efficient large language models. The two AI platforms, i.e., ChatGPT and DeepSeek, are among the most prominent AI platforms. They receive considerable attention in literature for their capacity to do several tasks: producing coherent text, optimizing work processes, and improving efficiency in several environments.

However, comparative studies have shown that both ChatGPT and DeepSeek have different yet specialized skills for specific tasks. In coding, DeepSeek outperformed the others in the correctness and optimality of codes, with ChatGPT making notable advances in suppressing syntax errors (Vyas & Bhardwaj, 2025). In business applications, ChatGPT effectively managed complex verbal interactions, whereas DeepSeek excelled in structured written content and grammar (AlAfnan, 2025). In the educational aspect, the use of both platforms was advised (Kotsis, 2025). ChatGPT was advised for energetic engagement and DeepSeek for planned studying (Kotsis, 2025). In medical applications, ChatGPT had higher accomplishments in standardized tests like USMLE, with DeepSeek being an apt free alternative (Zhang et al., 2025); both models are important in intricate diagnosis procedures as well (Liu et al., 2025). In learning languages, DeepSeek is superior in pinpointing contextual mistakes, with both models providing educational cues (Mahyoob & Algaraady, 2025). Invariably, ChatGPT is friendlier and more dynamic in interactions, with DeepSeek having greater transparency and technical superiority with special advantages in citation and data processing functions (Choi & Chang, 2025; Hajjo et al., 2025).

To the best of the researcher's knowledge, no study has yet applied Vinay & Darbelnet's model to analyze Arabic idiom translations generated by ChatGPT and DeepSeek. This study therefore conducts a comparative analysis to fill that gap. Furthermore, it evaluates the AI outputs for accuracy, cultural appropriateness, and naturalness using Vinay & Darbelnet's framework alongside Nida's theory of dynamic equivalence.

5.2 Nida's Dynamic Equivalence Theory

Eugene A. Nida's dynamic equivalence theory, later refined and renamed functional equivalence, remains one of the most influential theoretical frameworks in translation studies (Nida & Taber, 1969). The theory was developed particularly for Bible translation contexts

with the aim of achieving an equivalent effect on target readers rather than maintaining formal correspondence with source texts (Nida, 1964). Therefore, dynamic equivalence prioritizes naturalness and comprehensibility in the target language. Furthermore, the aim of the theory is to produce the same response in target readers as the original text produced in its initial audience. The theory has its supporters who argue that dynamic equivalence facilitates better cross-cultural communication as well as reader comprehension (Venuti, 1995). Accordingly, this is the rationale behind selecting this theory for the current study.

5.2.1 Previous Studies

Recent translation studies have applied Nida’s dynamic equivalence theory to specific texts (e.g., religious texts (See Thompson & Williams (2018) for Protestant Bible text, Chen & Nakamura (2020) for Buddhist text and Rahman (2021) for Islamic text translation, i.e., Hadith translation from Arabic to English), literary texts (e.g., Jiang (2020) for poetry translation, Nemati and Afzalifard (2017) for Children’s literature translation, Ziyi and Zheng (2023) for Novel translation), and technical and specialized texts (e.g., Huang & Wang (2024) for legal English text translation, Martinez et al. (2022) for Spanish-English medical document translation, Sufen (2019) for Business translation). Additionally, diverse language pairs have been examined via the application of Nida’s dynamic equivalence theory including English-Indonesian (Sari et al. , 2019; among others), Chinese-English (Li & Zhang, 2021), English-Urdu (Ahmad & Khan, 2020), Arabic-English (Al-Darraj, 2015; Alshammari, 2016; Jabak, 2020; Rahman, 2021), and European language combinations (Silva, 2019; Pietryga, 2020; among others). Furthermore, diverse methodological approaches have been employed by several recent studies for the analysis of dynamic equivalence application. These are corpus-based approaches (e.g., Rodriguez and Schmidt (2021), Sun and Wan (2022)), cognitive approaches (e.g., Sickinger (2017), Taylor and Johnson (2020)), and approaches using digital translation tools (e.g., Anderson et al. (2021), Wang and Li (2023)). Thus, this demonstrates both the theory’s continued relevance and its evolving methodological approaches in contemporary translation practice. However, no single study to date has applied Nida’s dynamic equivalence theory via artificial intelligence applications particularly ChatGPT and DeepSeek. Hence, the emergence of the importance of the current study which aims to fill in this research gap.

5.3 Vinay & Darbelnet’s Translation Strategies Theory

Vinay & Darbelnet’s translation strategies theory represents one of the foundational frameworks in translation studies. It was developed by French linguists Jean-Paul Vinay (1910-1999) and Jean Darbelnet (1904-1990). The theory was originally published in 1958 as “Stylistique comparée du français et de l’anglais” (Comparative Stylistics of French and English). Accordingly, this systematic approach emerged from their comparative analysis of French and English linguistic and stylistic features (Safi & Nasser, 2022). Then, it was adopted in other languages by different researchers (Munday, 2016, p. 88)

The theory provides a comprehensive framework for understanding translation processes. This is achieved through seven distinct procedures organized into two main strategies: direct translation and oblique translation. Direct Translation includes three procedures. These are: (1) Borrowing, (2) Calque, and (3) Literal Translation. On the other hand, Oblique Translation includes four procedures. These are: transposition, modulation, equivalence, and adaptation. Furthermore, these procedures operate across three analytical levels. These levels are: “the lexicon” (i.e., examining the repertoire of signs through substitution of translation units), “syntactic structure” (i.e., focusing on the arrangement of translation units in sequence), and the message level (i.e., addressing individual entities where speakers determine point of view through tone, register choice, and textual organization) (El-Farhaty, 2015, p. 59; Albogdady, 2023).

The below table (Table 1) summarizes the meaning of each procedure and its aim, and it gives examples for the readers.

Table 1. Summary of Seven Procedures

Direct Translation Procedures	Meaning	Aim	Examples
1. Borrowing	A translation procedure where a word (or expression) is taken directly from the source language without any modification (Vinay & Darbelnet, 1995). The borrowed element is integrated into the target language text in its original form. This is often due to the absence of an equivalent term in the target language.	The aim is (1) to fill lexical gaps in the target language when no equivalent exists or (2) to preserve the cultural specificity and authenticity of certain concepts that are unique to the source culture (Munday, 2016). This procedure maintains the original cultural connotations and prevents loss of meaning through inadequate translation.	Arabic to English: “الحج” (al-hajj) → “hajj”; “الإمام” (al-imam) → “imam”; “الخليفة” (al-khalifa) → “caliph”. English to Arabic: “internet” → “إنترنت”; “computer” → “كمبيوتر”.
2. Calque	It involves the literal translation of a foreign word (or phrase) by translating each of its elements separately. Thus, it creates a new expression in the target language that mirrors the structure of the source language (Vinay & Darbelnet, 1995).	The aim is to create new expressions in the target language that preserve the semantic as well as structural relationships of the source language while making the concept accessible to target language speakers (Baker, 2018). This allows for the introduction of foreign concepts while maintaining linguistic naturalness.	English to Arabic: “United Nations” → “الأمم المتحدة” (al-umam al-muttahida - literally “the nations the united”); Arabic to English: “بيت المال” (bayt al-mal) → “house of money”

3. Literal Translation	It is a translation procedure where the source language text is translated word-for-word into the target language. This maintains the same syntactic structure and semantic relationships, and is considered acceptable when the result is both grammatically correct and idiomatically appropriate in the target language (Vinay & Darbelnet, 1995).	The aim is to preserve the exact meaning and structure of the source text when the linguistic systems of both languages allow for direct correspondence (Hatim & Munday, 2019). This procedure ensures maximum fidelity to the original while producing natural-sounding target language text.	English to Arabic: "I love you" → "أنا أحبك" (ana uhibbuk)- Arabic to English: "السلام عليكم" (as-salamu 'alaykum) → "Peace be upon you"; "الحمد لله" (al-hamdu lillah) → "Praise be to God".
Indirect Translation Procedures	Meaning	Aim	Examples
4. Transposition	It is a translation procedure that involves changing the grammatical category or word class of elements in the source text while preserving the meaning, such as changing a noun to a verb, or altering the syntactic structure of a sentence (Vinay & Darbelnet, 1995). This procedure addresses structural differences between languages.	The aim is to produce natural and idiomatic expressions in the target language when direct structural correspondence is not possible or would result in awkward constructions (Newmark, 1988). This procedure ensures grammatical correctness as well as stylistic appropriateness in the target language.	English to Arabic: "No smoking" (noun) → "ممنوع التدخين" (mamnu' at-tadkhin - "forbidden is smoking" - verb form); Arabic to English: "بعد وصوله" (ba'd wusulih - "after his arrival" - noun) → "After he arrived" (verb)
5. Modulation	A translation procedure that involves changing the point of view, perspective, or cognitive category of the source language message; At the same time, it maintains the same meaning, often it involves shifts in semantic focus or conceptual framing (Vinay & Darbelnet, 1995).	The aim is to achieve naturalness and idiomaticity in the target language. This is created by adapting the expression to the thought patterns and conventional ways of expression typical of the target culture and language (Shuttleworth & Cowie, 2014). This ensures cultural as well as linguistic appropriateness.	English to Arabic: "You are wrong" → "أنت مخطئ" or "لست محقاً" (lasta muhiqan - "you are not right"); Arabic to English: "ما شاء الله" (ma sha' Allah - literally "what God willed") → "How wonderful!" or "Amazing!".
6. Equivalence	A translation procedure used to render the same situation using completely different stylistic and structural means. This procedure is typically applied to idiomatic expressions, proverbs, and culturally-bound phrases where word-for-word translation is impossible (Vinay & Darbelnet, 1995).	The aim is to reproduce the same communicative effect and pragmatic function in the target language. This is applied when dealing with expressions that have no direct linguistic correspondence but serve equivalent communicative purposes in their respective cultures (Pym, 2023).	English to Arabic: "It's raining cats and dogs" → "تمطر بغزارة" (tumtir bi-ghazara - "it's raining heavily"); Arabic to English: "على قد لحافك مد رجلك" (literally "stretch your legs according to your blanket") → "Cut your coat according to your cloth".
7. Adaptation	It is a translation procedure that involves the highest degree of departure from the source text. Thus, the translator replaces cultural references, situations, or concepts from the source culture with equivalent ones from the target culture. The aim is to ensure comprehensibility and relevance for the target audience (Vinay & Darbelnet, 1995).	The aim is to (1) achieve cultural equivalence and (2) ensure that the translated text resonates with the target audience by replacing culture-specific elements that would be unfamiliar or meaningless to target readers with culturally appropriate alternatives (Bassnett, 2013).	English to Arabic: "Christmas dinner" → "وليمة العيد" (walima al-'id - "feast of celebration") Arabic to English: "صلاة الجمعة" (salat al-jum'a) → "Sunday service" (in Christian context).

5.3.1 Recent Studies Employing Vinay & Darbelnet’s Theory and Significance of The Model

Vinay & Darbelnet’s model has been applied across diverse translation contexts. It has been applied to literary translation context (See, for example, these studies: Dwipuspita and Wediyantoro, 2023; Hassan and Hussein, 2024; Tavakoli and Zare, 2024; Suprayetno et al., 2024), religious and cultural contexts (Al-Fdhl, 2023; Hani and Zainab, 2024), dramatic and media contexts (Aziz, 2022; Al-Mizgagi and Al-Tamemi, 2023), comparative and analytical contexts (Kocabiyik, 2022; Albogdady, 2023). Thus, one can say that Vinay & Darbelnet’s translation strategies theory remains a cornerstone of translation studies. Furthermore, it demonstrates remarkable longevity and adaptability since its 1958 inception. The significance of the model lies in several points: (1) its systematic approach (as it provides a systematic as well as comprehensive framework that moves translation analysis beyond intuitive approaches toward empirical methodology (Safi & Nasser, 2022)), (2) its cross-cultural applicability (as recent studies spanning Middle Eastern, South Asian, and European contexts indicate the framework’s cross-cultural applicability while revealing culture-specific translation patterns (Hassan & Hussein, 2024; Tavakoli & Zare, 2024; Suprayetno et al., 2024)), and its (3) pedagogical value (as the model helps students understand translation as a systematic process requiring strategic decision-making rather than mere linguistic substitution and this is crucial for translation education). Thus, these abovementioned points ensure the model ongoing significance in contemporary translation studies as well as practice. One can say that the enduring importance of the theory lies not only in its comprehensive analytical framework but also in its ability to bridge theoretical understanding with practical application. Thus, this model supports translators, educators, and researchers in the field of translation studies as it is an indispensable tool in their hands. Despite its application in various contexts (as explained above), as well as its significance, Vinay & Darbelnet’s model was not applied in relation to artificial intelligence (AI) context as well idioms. To the best of the researcher knowledge, no single study to date has applied Vinay & Darbelnet’s model to the translation of idioms given by ChatGPT and DeepSeek. Thus, the present study aims to fill in this gap by conducting a comparative analysis of Arabic idiom translations generated by ChatGPT and

DeepSeek. The aim is to evaluate ChatGPT and DeepSeek accuracy, cultural appropriateness, and naturalness using Vinay & Darbelnet's model along other established translation frameworks (i.e., Nida's theory of dynamic equivalence).

5.4 Neural Machine Translation and AI Handling of Idiomatic and Cultural Language

Machine Translation affects translation in various ways. It influences both the process and the product (Jibreel, 2023). Furthermore, Neural machine translation (NMT) and large language models (LLMs) like ChatGPT and DeepSeek have enhanced automated translation. This was created by leveraging advancements in machine learning techniques like encoder-decoder models and attention mechanisms. Accordingly, they allow for contextualized language processing (Tan et al., 2020). These models were trained on large multilingual datasets and, therefore, they produce fluent translations (Baziotis et al., 2023). However, they perform poorly two points: culturally nuanced and idiomatic expressions, because these expressions require a deeper cultural understanding beyond statistical patterns (Haider et al., 2025; Almahasees, 2021). The complexity emerges in the balance between formal equivalence (literal translations) and dynamic equivalence (culturally appropriate translations). Idioms in the current study, for instance, would require dynamic equivalence and, therefore, the AI may struggle with phrases like "kick the bucket" in translation, producing awkward translation (Vinay & Darbelnet, 1995; Aalfaqih, 2025). This is most especially evident in languages like Arabic. The reason for that is its owing on culturally specific metaphors and proverbs (Farghal & Haider, 2024). Studies have also reviewed effective prompt designs (e.g., Chan & Tang, 2024; Cao et al., 2025, among others). The quality of GPT-generated translations heavily relies on effective prompt design (Chan & Tang, 2024). When specific details were provided such as purpose, target audience, task information, pre-edit schemes, and examples, translations improved significantly (Chan & Tang, 2024). This leads to nuance prompt engineering being a critical aspect of AI-assisted translation which aids in producing the desired quality (Cao et al., 2025). Future advancements aim to enhance AI's cultural understanding through techniques like retrieval augmentation and multimodal learning, but true cultural fluency will likely require hybrid human-AI collaboration (Liu et al., 2023).

The current study fills gaps in literature by addressing several points. These are the lack of (1) comparative studies, (2) evaluation frameworks, and (3) post-editing guidelines for idiomatic translations. Furthermore, it aims to enhance machine translation technologies by offering insights into handling idiomatic expressions with cultural as well as contextual understanding.

6. Method of Analysis

6.1 Theoretical Framework

The theoretical foundation for this study lies in the intersection of translation studies as well as computational linguistics. It focuses on the theories of equivalence in translation and natural language processing. Nida's theory of dynamic equivalence provides a guiding principle for evaluating how well idioms are rendered in target languages—not just through literal accuracy, but through meaning as well as cultural impact. This is supplemented by Vinay and Darbelnet's translation strategies. It highlights the importance of specific strategies such as transposition and modulation when dealing with idioms.

From a computational perspective, this study draws on concepts from neural machine translation (NMT). The focus is on how large language models interpret and generate contextually appropriate language. The performance of ChatGPT and Deep Seek is analyzed in light of these theories, particularly how machine learning models manage semantic equivalence, register, and figurative language. By integrating insights from both theoretical and technological frameworks, this paper evaluates how closely AI-generated translations align with human expectations of idiomatic fidelity and cultural relevance.

6.2 Data Collection and Procedures

The current study compares the Arabic translations of 20 English idioms generated by ChatGPT and Deep Seek. These (English) idioms belong to a category which is entitled by Ghazala (2008) as "direct idioms" (p.128). The reason for considering them direct idioms is that they are translated directly and at the same time should be understood indirectly. This is because they have metaphorical meanings. To give an example, the intended direct meaning of the idiom 'passing the exam is not a bed of roses' (النجاح في الامتحان ليس طريقا مفرشا) (بالورود) is 'passing the exam is difficult' (i.e., النجاح في الامتحان أمر صعب). The selected idioms were numbered from one to twenty (See Appendix A entitled Data). Each idiom was translated by both tools without additional context (See Appendix B for figures of each tool, i.e., figures from 1 to 4). The date of submission of idioms to both DeepSeek and ChatGPT is 24/3/2025 (as shown in Appendix B, Figure 5). Translations were then evaluated by bilingual language experts using three criteria. These are (1) accuracy (Evaluates whether the meaning of the idiom is correctly conveyed in Arabic.), (2) cultural appropriateness (Assesses whether the translation is in line with Arabic linguistic norms and culturally equivalent idioms), and (3) naturalness (Judges how fluent and readable the translation is for native Arabic speakers.) (See Appendix C, Table 13). Then, two theoretical frameworks from translation studies— (1) Nida's dynamic equivalence and (2) Vinay and Darbelnet's translation strategies—alongside principles of neural machine translation (NMT) were applied (See tables from 3 to 12 in data analysis, results and discussion section, and tables from 5 to 29 in Appendix E). The evaluation process involved qualitative analysis and scoring.

7. Data Analysis, Results and Discussion

7.1 Tripartite Analysis: Accuracy, Cultural Appropriateness, and Naturalness

The translations generated by ChatGPT and DeepSeek are given in table 13 in Appendix C. Furthermore, the table provides a quantitative evaluation of the translations, across 20 English-Arabic idiom pairs (See Table 13: quantitative evaluation of translations generated by ChatGPT and DeepSeek in Appendix C). Each translation is assessed against three criteria: Accuracy (i.e., correct meaning conveyance),

Cultural Appropriateness (i.e., alignment with Arabic linguistic norms), and Naturalness (i.e., fluency for a native speaker). Points are awarded per example: 1 point for a superior rating, 0.5 points to each system for a tie, and 0 points for an inferior rating. The points from all 20 examples were tallied for each system and criterion. The maximum possible score per system is 20 points per criterion (1 point per example). Thus, the below table (Table 2) gives the readers the total points for both ChatGPT and DeepSeek.

Table 2. Performance of ChatGPT and DeepSeek

Criterion	ChatGPT Total Points	DeepSeek Total Points
Accuracy	14.5 / 20	15.5 / 20
Cultural Appropriateness	14.0 / 20	18.0 / 20
Naturalness	14.0 / 20	17.0 / 20
GRAND TOTAL	42.5 / 60	50.5 / 60

The Percentage Scores for ChatGPT is 70.8%, and for DeepSeek is 84.2%

The results clearly indicate that DeepSeek outperforms ChatGPT in the translation of English idioms into Arabic, with a significant 13.4% overall point differential. Furthermore, the table holds a comparison between the two tools that covers key metrics such as accuracy, cultural appropriateness, and naturalness of the translations. Obviously, DeepSeek shows a consistent advantage in all areas. In terms of accuracy, both systems demonstrate high competency in conveying core meaning. ChatGPT achieved a score of 72.5%, while DeepSeek scored slightly higher at 77.5%. The minor advantage for DeepSeek stems from its ability to handle metaphors more effectively. For example, in Examples 1 and 12, DeepSeek provided additional idiomatic translations that preserved the figurative meaning, which ChatGPT either omitted or translated too literally. (See Table 13 in Appendix C).

Cultural appropriateness is the area with the most significant difference between the two systems. ChatGPT scored 70%, while DeepSeek excelled with a score of 90%. DeepSeek’s translation strategy consistently favors cultural adaptation, as seen in translations like المصالح المتبادلة in Example 3, تنشر مشاكلك in Example 12, and يسبح في الأموال in Example 14. ChatGPT, although often accurate, tends to default more frequently to literal translations (or generalizations), which, while understandable, lack the cultural resonance of a native Arabic idiom. This is the primary driver of DeepSeek’s superior performance overall. Naturalness is another area where DeepSeek outperforms ChatGPT, scoring 85% compared to ChatGPT’s 70%. This metric is closely tied to cultural appropriateness, as DeepSeek’s translations are often rated as more fluent and natural. By employing common Arabic idioms, DeepSeek’s output reads less like a translation and more like original content, i.e., Arabic. ChatGPT’s translations, while grammatically sound, tend to have a “translated” feel, being flatter and less evocative. Examples 1, 3, and 14 illustrate how DeepSeek’s translations feel more integrated into the target language, while ChatGPT’s translations appear more mechanical.

There are, however, notable outliers where ChatGPT outperformed DeepSeek in specific cases. For instance, in Example 10, ChatGPT provided a more concise translation, and in Example 5, it produced a more natural word order. These examples show that while DeepSeek excels in idiomatic translation, ChatGPT can outperform it in certain, typically non-idiomatic, syntactic choices (See examples 5 and 10 in tables 17 and 21 respectively in appendix E). Example 9, (See tables 7 and 8), was a failure for both systems in terms of cultural appropriateness and naturalness. This highlights a shared limitation for both tools. Thus, the role of the translator is important in critically evaluating the generated translations and modifying them to suits Arabic readers.

In conclusion, this point-based analysis confirms that while both NMT systems are powerful tools, DeepSeek provides a superior translation experience for idioms. Its algorithms seem more optimized for identifying and utilizing dynamic equivalence. Furthermore, it prioritizes meaning and cultural impact over literal wording. ChatGPT remains a highly capable engine but is more likely to produce a translation that requires post-editing to achieve a truly natural and idiomatic quality in the target language.

7.2 Nida’s Dynamic Equivalence, Vinay & Darbelnet’s Strategies, and NMT Performance

This section presents a comprehensive analysis of data, results, and discussions concerning idiomatic expressions that highlight the differences between the two MNT systems, DeepSeek and ChatGPT, in their translations. The analysis is framed within the theoretical frameworks of Nida’s Dynamic Equivalence, Vinay and Darbelnet’s Translation Strategies, and NMT performance. The specific idioms examined are 1, 3, 9, 12, and 14. The remaining idiomatic expressions are provided in tabular form, with each entry detailed in Appendix E.

Table 3. Arabic Translations of Idiom 1: “Passing the exam is not a bed of roses.”

ChatGPT Translation	DeepSeek Translation
اجتياز الامتحان ليس بالأمر السهل	اجتياز الامتحان ليس سهلاً (أو: ليس كالسير على الورود)

Table 4. Analysis of Example 1: “Passing the exam is not a bed of roses.”

Framework	Analysis
Nida’s Dynamic Equivalence	The expression ‘a bed of roses’ in English means a situation or activity that is comfortable or easy (Mounassar, 2024, p. 100). ChatGPT’s Translation (ليس بالأمر السهل) achieves only formal equivalence. It renders the basic message (denotation) but completely fails to reproduce the rhetorical effect of the idiom. Thus, the target reader receives the information but does not experience the same vivid, metaphorical understanding as the source reader. DeepSeek’s Option 2 (ليس كالسير على الورود) achieves a much higher degree of dynamic equivalence. It prioritizes the <i>effect</i> over the <i>form</i> . By changing the metaphor from a “bed” to “walking on,” it finds a natural and culturally resonant way to evoke the same feeling—that the task is far from pleasant or simple. The reader’s response is aligned with the original intent. Furthermore, DeepSeek’s Option 2 is close to the translation given by Ghazala (2008) “ليس طريقاً مفروشاً بالورود” (p. 128).
Vinay & Darbelnet’s Strategies	ChatGPT uses a strategy of Generalization (or Modulation). It reduces the specific, image-rich idiom to a broader, more general concept (“not easy”). This is a safe but unambitious strategy that loses nuance. DeepSeek demonstrates a more sophisticated approach. For Option 2, it uses Equivalence (for the idiomatic meaning) combined with Adaptation/Transposition. It adapts the original metaphor (“bed”) to one that works better in Arabic (“walking”), showing an understanding that a direct literal translation (سير الورود) would be nonsensical. This is a creative and effective solution to a common translation problem. One can say that DeepSeek’s Option 2 conveys the intended meaning by employing an idiomatic expression which is familiar to Arabic speakers.
NMT Performance	DeepSeek’s performance is notably superior in this idiom. The key differentiator is its ability to offer multiple translation strategies: (1) Providing Option 1 (ليس سهلاً) serves the user who needs a quick, clear meaning and (2) providing Option 2 (ليس كالسير على الورود) serves the user who needs a stylistically faithful, idiomatic translation. This multi-output approach demonstrates an advanced, context-aware NMT model that understands translation is not about finding one “correct” answer but about making choices based on context, purpose, and desired effect. ChatGPT’s single, generalized output is functional but less helpful for a user seeking to understand the cultural nuance of the original phrase (which is an idiomatic expression).

In conclusion, the analysis reveals a significant distinction in the approaches taken by ChatGPT and DeepSeek in translating the idiomatic expression “Passing the exam is not a bed of roses.” ChatGPT’s translation, while accurate in denotation, fails to capture the vivid, metaphorical nuance of the original expression. Thus, it offers only a formal equivalence that may not resonate with the target audience. In contrast, DeepSeek demonstrates a higher level of dynamic equivalence. This is achieved by effectively adapting the metaphor to one that aligns with Arabic cultural understanding. This creative and context-aware approach not only preserves the meaning but also evokes the same emotional response as the source text. Furthermore, DeepSeek’s use of multiple translation options: one literal and one idiomatic, highlights its sophisticated, context-driven strategy. Therefore, it better addresses the varied needs of the target audience. Ultimately, DeepSeek outperforms ChatGPT in this instance. It showcases the importance of flexibility and cultural adaptation in achieving a faithful and effective translation.

Table 5. Arabic Translations of Idiom 3: “You scratch my back and I scratch yours.”

ChatGPT Translation	DeepSeek Translation
قدم لي خدمة لأردھا لك	أنت تساعدني وأساعذك (أو: المصالح متبادلة)

Table 6. Analysis of Example 3: “You scratch my back and I scratch yours.”

Framework	Analysis
Nida’s Dynamic Equivalence	ChatGPT’s Translation (قدم لي خدمة...) operates at the level of formal correspondence. It translates the words and the grammatical structure but fails at reproducing the pragmatic effect. The original is a well-known, slightly informal idiom implying a mutual understanding. DeepSeek’s Option 2 (المصالح متبادلة) can be considered as an example of achieving dynamic equivalence. The reason is that the meaning of the translation is “our interests are aligned for mutual benefit.” DeepSeek uses the natural Arabic expression. It produces the same effect on an Arabic speaker. The form is different, but the response evoked is identical. This is very close to the Arabic translation given by Ghazala (2008) كل وأكل-حكلي لحكلك -امسكلي واقطعك أضي لي اقدح لك- for the idiom “You scratch my back and I scratch yours.”
Vinay & Darbelnet’s Strategies	ChatGPT uses a combination of Literal Translation and Generalization. It attempts to translate the words “scratch” and “back” conceptually (خدمة- service) but remains too close to the original structure. This leads to calque. DeepSeek employs the Equivalence (or Adaptation) strategy in its translation. It recognizes that a literal translation is impossible and would be nonsensical. Instead, it jettisons the original imagery entirely and selects a pre-existing cultural equivalent in the target language that serves the same communicative purpose. This is a high-level translation strategy that demonstrates deep cultural understanding
NMT Performance	This example clearly demonstrates DeepSeek’s superior performance. Its ability to generate “المصالح متبادلة” shows it has been trained on a robust corpus of parallel texts where this equivalence is established. It doesn’t just translate words; it translates meaning and usage. ChatGPT’s output reveals a more simplistic model that often defaults to literal or generalized paraphrasing. It struggles with culturally specific phrases. Obviously, it translates the semantic meaning. However, it does not give the pragmatic function. Furthermore, DeepSeek’s offering of two options (Option 1: a direct paraphrase; Option 2: a cultural idiom) provides the user with choice and context, which is a hallmark of a more advanced and user-centric NMT system. This shows the point that the best translation depends on the desired style as well as nuance.

In conclusion, the analysis of ChatGPT and DeepSeek generated translations reflects the importance of cultural as well as pragmatic adaptation in the process of translation. ChatGPT provides a literal translation that may convey the basic meaning. But it fails to capture the idiomatic essence as well as the natural flow of the original phrase. On the other hand, DeepSeek excels by employing a dynamic equivalence approach. Thus, it offers a translation that resonates with Arabic speakers through the culturally appropriate expression “المصالح متبادلة.” It conveys not only the same meaning but the same effect as the English idiom as well. This indicates a deep understanding of both languages’ cultural nuances. Furthermore, the use of multiple options (on the part of DeepSeek) gives users flexibility based on their specific needs. Whether they want a direct paraphrase (by selecting option 1) or a culturally rich equivalent (by selecting option 2). This multi-output approach and attention to context make DeepSeek a more advanced and user-centered NMT system if compared to ChatGPT (which provides literal translation style).

Table 7. Arabic Translations of Idiom 9: “They signed on the dotted line.”

ChatGPT Translation	DeepSeek Translation
وقعوا على العقد دون تردد	وقعوا على الخط المنقط

Table 8. Analysis of Example 9: “They signed on the dotted line.”

Framework	Analysis
Nida’s Dynamic Equivalence	ChatGPT’s Translation is a clear failure in dynamic equivalence. It does not produce the same effect on the target reader as the original does on the source reader. The original evokes the final, formal act of agreement. Siefring (2004) refers to the idiom’s meaning: “agree formally” (p. 262). The translation generated by ChatGPT evokes a sense of eagerness (or lack of doubt). It translates a perceived implication instead of the actual message. DeepSeek’s Translation also fails to achieve full dynamic equivalence. While it correctly identifies the words, the effect is wrong. The original idiom feels natural and conclusive to an English speaker. The Arabic calque feels technical, awkward, and foreignizing. The reader understands the mechanics but not the natural idiom. True dynamic equivalence would require finding an Arabic expression that conveys the sense of finalizing a formal commitment, even if it doesn’t mention signing or lines. For example, أبرموا الصفقة (They closed the deal) or وافقوا بشكل نهائي (They agreed finally) or “وقعوا على بياض” as given by Ghazala (2008, p. 129) would be closer in spirit.
Vinay &	ChatGPT employs Generalization combined with an erroneous Amplification (adding “without hesitation”). It abandons the

Framework	Analysis
Darbelnet's Strategies	specific imagery of the idiom altogether in favor of a broader, and in this case incorrect, meaning. DeepSeek uses a strict Literal Translation (or Calque). This strategy is ineffective here because the literal components ("dotted line") do not assemble into a recognized idiom in the target language. The result is a foreign-sounding phrase. The optimal strategy here would be Modulation (change of point of view) or Adaptation. This involves moving away from the literal image to express the same underlying reality in a natural way for the target culture. A good translation would drop the "dotted line" entirely and focus on the concept of formal agreement.
NMT Performance	This example highlights a classic challenge for NMT systems: translating idioms that are deeply cultural and non-literal. ChatGPT's error is a common one in machine translation: it relies on statistical patterns. It may have seen "signed on the dotted line" correlated with contexts implying eagerness or commitment and generated a frequent, related phrase (دون تردد), resulting in a critical loss of meaning. DeepSeek's approach is the more conservative and technically "safer" NMT strategy (i.e., it translates the words it see). Obviously, this technique avoids errors but the result is unidiomatic output. Therefore, it requires human post-editing.

In conclusion, both systems struggle with the balance between literal translation and cultural adaptation. This underscores the difficulty of translating idiomatic expressions without a deep contextual understanding. For true dynamic equivalence, a more adaptive approach is needed. This is the one that focuses on conveying the underlying concept of formal agreement in a culturally resonant way. Thus, it is recommended to select Ghazala's translation (2008) of this idiom as "وقعوا على بياض" (p. 128). Ultimately, the analysis of generated translations of this idiom highlights the limitations of current NMT systems in handling idioms that require not only linguistic accuracy but also cultural sensitivity.

Table 9. Arabic Translations of Idiom 12: "Why do you wash your dirty linen in public?"

ChatGPT Translation	DeepSeek Translation
لماذا تغسل غسيلك القذر أمام الناس؟	لماذا تنتشر مشاكلك أمام الجميع؟

Table 10. Analysis of Example 12: "Why do you wash your dirty linen in public?"

Framework	Analysis
Nida's Dynamic Equivalence	ChatGPT's Translation fails to achieve dynamic equivalence. It prioritizes formal correspondence (word-for-word translation), resulting in an output that may be decoded by the target audience but will not evoke the same natural, idiomatic response. The reader is likely to be confused by the bizarre imagery of publicly washing literal dirty clothes. DeepSeek's Translation is an example of successful dynamic equivalence. It asks: "What is the effect of this idiom on an English speaker?" The answer is: "It criticizes someone for revealing private, potentially shameful, matters in public." DeepSeek then finds the natural Arabic expression (تنتشر مشاكلك) that produces the exact same effect on an Arabic speaker. Obviously, the form is completely different, but the meaning and impact are identical. Ghazala (2008) translates the above-mentioned idiom into Arabic and uses the same verb "تنتشر" (p. 129) (which is given by DeepSeek's Translation): "لماذا تنتشر غسيلك الوسخ على الملا؟". Thus, DeepSeek's translation is very close to the translation of Gazala (2008).
Vinay & Darbelnet's Strategies	ChatGPT uses a Literal Translation strategy. This usage of literal translation ignores the cultural embeddedness of the expression. Thus, it is considered as the least effective strategy for idioms. On the other hand, DeepSeek employs the Equivalence (or Adaptation) strategy. It recognizes that the source-language idiom has no literal equivalent in the target language. Therefore, it abandons the original metaphor entirely and selects a pre-existing cultural analogue that serves the same communicative purpose. This is a high-level, creative translation technique that demonstrates deep cultural fluency.
NMT Performance	This example indicates a contrast in NMT sophistication. ChatGPT's performance is characteristic of a simpler model that relies on direct lexical mapping. It translates the surface-level words without comprehending the underlying idiomatic meaning. This leads to a flawed output. This supports the results given in Almahasees (2021) and Haider et al. (2025) regarding the poor points of AI: culturally nuanced and idiomatic expressions. DeepSeek's performance indicates a more advanced model. Its ability to generate "تنتشر مشاكلك" suggests it has been trained on a vast corpus of parallel texts where this equivalence is established.

In conclusion, this analysis demonstrates a clear distinction in the translation approaches of ChatGPT and DeepSeek. ChatGPT's literal translation fails to capture the cultural as well as the contextual meaning behind the idiom. On the other hand, DeepSeek successfully applies dynamic equivalence. This achieved by shifting away from the original metaphor and using a culturally relevant expression ("تنتشر")

مشاكل). This usage evokes the same effect in the target language. Furthermore, the translation aligns with the original intent of the idiom and is supported by the translation provided by Gazala (2008).

Table 11. Arabic Translations of Idiom 14: “That man is rolling in money.”

ChatGPT Translation	DeepSeek Translation
ذلك الرجل يغرق في المال	ذلك الرجل يسبح في الأموال

Table 12. Analysis of Example 14: “That man is rolling in money.”

Framework	Analysis
Nida’s Dynamic Equivalence	While both translations convey the core meaning of extreme wealth, they do not achieve dynamic equivalence to the same degree. The key question for Nida is: “What is the response of the target language reader?” For يسبح في الأموال (DeepSeek), the response is more closely aligned with that of an English speaker hearing “rolling in money”: an immediate understanding that the person is extremely rich, coupled with a vivid, natural, and idiomatic metaphorical image. The translation يغرق في المال (generated by ChatGPT), though accurate, produces a slightly “flatter” response, as the metaphor is less common and feels less instinctive. DeepSeek’s version better prioritizes the natural reader experience.
Vinay & Darbelnet’s Strategies	Both models correctly employ the strategy of Modulation (a change of point of view). However, they differ in their choice of metaphor-to-metaphor substitution. DeepSeek’s selection of يسبح (to swim) demonstrates a more culturally appropriate and idiomatic adaptation of the concept of being completely surrounded by money. ChatGPT’s choice of يغرق (to drown), while conceptually similar, represents a less common and therefore less optimal modulation within the target culture. Ghazala (2008) refers to the meaning of this idiom: “ذاك الرجل ثري جداً” (p. 130)
NMT Performance	Obviously, both systems successfully moved beyond literal translation. DeepSeek managed to select the most natural and culturally resonant equivalent. The variation in output (يسبح vs. يغرق) is not just a matter of stylistic choice. It reflects a difference in the quality of cultural-linguistic understanding. Thus, DeepSeek’s translation indicates training on a corpus that more accurately reflects contemporary, idiomatic usage. Therefore, it allows it to make a more nuanced and appropriate lexical selection.

In conclusion, the results of this analysis show that ChatGPT and DeepSeek achieved varying degrees of dynamic equivalence when translating the idiom. Despite the fact that both translations convey the core meaning, DeepSeek’s choice of “يسبح في الأموال” is more idiomatic and culturally resonant. It evokes a vivid and natural metaphor that closely matches the response of an English speaker to hearing “rolling in money.” In contrast, ChatGPT’s “يغرق في المال” is accurate but less intuitive because the metaphor of drowning is less common and feels less immediate in this setting.

8. Conclusion

This study aimed to conduct a comparative analysis of Arabic idiom translations generated by ChatGPT and DeepSeek. It has focused on three main evaluation criteria: accuracy, cultural appropriateness, and naturalness. Employing established translation frameworks, i.e., Nida’s Dynamic Equivalence Theory and Vinay & Darbelnet’s translation strategies, the study sought to evaluate how effectively these two neural machine translation (NMT) systems handled idiomatic expressions and culturally embedded phrases. Based on the analysis of 20 idioms examples, the study provides insights into the strengths and limitations of both translation tools in the task of idiomatic translation between English and Arabic.

In response to the first research question: ‘How accurately can AI tools translate English idioms into Arabic while preserving their intended meaning?’, the study reached the conclusion that both ChatGPT and DeepSeek demonstrated a high level of accuracy in translating idiomatic expressions, with slight differences in their ability to preserve meaning. For example, when translating idioms like “He killed his neighbour in cold blood” (قتل جاره بدم بارد) or “Their company is on the black list” (شركتهم مدرجة في القائمة السوداء), both systems accurately conveyed the literal meaning of the expressions. However, ChatGPT sometimes failed to maintain the metaphorical richness of certain idioms, such as in the translation of “Passing the exam is not a bed of roses”. ChatGPT’s literal translation “الاجتياز الامتحان ليس بالأمر السهل” (Passing the exam is not easy) missed the figurative nature of the English expression, which DeepSeek preserved more effectively with “ليس كالسير على الورود” (Not like walking on roses). This highlights DeepSeek’s slight edge in accurately translating idioms that rely on metaphorical language. Thus, while both systems performed well in conveying the core meaning of the idioms, DeepSeek demonstrated a superior ability to capture the full metaphorical essence of idiomatic expressions, suggesting it may be better suited for translating idiomatic and figurative language. ChatGPT, on the other hand, excelled at literal accuracy, though sometimes at the expense of losing idiomatic richness.

Regarding the second research question, to what extent do AI translations reflect culturally appropriate and idiomatically equivalent expressions in Arabic? which examines the cultural appropriateness of the generated translations, the study reached the conclusion that DeepSeek consistently outperformed ChatGPT. For example, when translating “*You scratch my back and I scratch yours*”, ChatGPT offered the translation “*قدم لي خدمة لأردها لك*” (Do me a favor, and I will return it), which, while accurate, lacks the idiomatic flavor of the original expression. DeepSeek, however, provided the translation “*أنت تساعدني وأساعدك (أو: المصالح متبادلة)*” (You help me, and I help you / Mutual interests), which is culturally more fitting and idiomatic in Arabic. Similarly, when translating “*Why do you wash your dirty linen in public?*”, DeepSeek adapted the phrase into “*لماذا تنتشر مشاكلك أمام الجميع؟*” (Why air your problems in public?), which is a more culturally appropriate expression in Arabic, while ChatGPT provided a more literal translation, “*لماذا تغسل غسيلك القذر أمام الناس؟*” (Why do you wash your dirty laundry in front of people?), which is less idiomatic in Arabic culture. Thus, these examples illustrate that DeepSeek’s translations are more culturally sensitive, providing expressions that resonate with native Arabic speakers. ChatGPT, while often accurate, leaned more towards literal translations, which could lead to a loss of cultural resonance in certain idiomatic contexts. The overall cultural appropriateness of DeepSeek’s translations, as seen in its use of expressions like “*المصالح متبادلة*” (Mutual interests) and “*انتشر مشاكلك*” (Air your problems), significantly contributed to its higher performance in this criterion.

As for the third question, ‘How natural and fluent are the Arabic translations produced by AI tools from the perspective of native speakers?’, which is related to examining naturalness of the translations, the results of the study showed that DeepSeek consistently provided more fluent and idiomatic translations compared to ChatGPT. In expressions like “*You make my blood boil*” or “*That man is rolling in money*”, DeepSeek produced translations that read like native Arabic idioms. For example, “*ذلك الرجل يسبح في الأموال*” (That man swims in wealth) felt natural and culturally appropriate, whereas ChatGPT’s “*ذلك الرجل يغرق في المال*” (That man is drowning in money) was accurate but felt slightly flat and less idiomatic. Similarly, when translating “*You make my blood boil*”, ChatGPT produced “*تصرفاتك السيئة تجعل دمي يغلي*” (Your bad actions make my blood boil), which was accurate but lacked the fluent feel that DeepSeek’s translation, “*تغلي دمي*” (Your bad actions make my blood boil), brought. In some cases, ChatGPT did provide more natural-sounding translations, such as in “*She cannot believe her eyes/ears*” (“*لا تصدق ما تراه/تسمعه*”) where it was more concise and fluent than DeepSeek, whose translation “*لا تصدق ما تراه عيناها/تسمعه أذناها*” was slightly more formal and redundant. Nevertheless, these instances were outliers, and overall, DeepSeek’s translations were rated as more natural and fluent due to its frequent use of common Arabic idioms.

Regarding the fourth question: What are the comparative strengths and limitations of ChatGPT and DeepSeek in the idiomatic translation task?, the study reached the conclusion that several key differences emerge. ChatGPT’s strengths lie in its literal accuracy and conciseness. It often provides translations that preserve the exact meaning of the original idiom. However, it struggles with idiomatic richness, and its translations sometimes feel flat or lack cultural resonance. For example, ChatGPT was less effective in idioms like “*You scratch my back and I scratch yours*” or “*Passing the exam is not a bed of roses*”, where it provided more literal translations without adapting them to the cultural or idiomatic context of the target language.

9. Recommendations of the Study

Based on the comparative analysis of the 20 English idioms translated by ChatGPT and DeepSeek, several key recommendations emerge for (1) improving machine translation (MT) systems, (2) enhancing translator training, and (3) guiding future research. The analysis highlighted both strengths and weaknesses in terms of accuracy, cultural appropriateness, and naturalness.

For AI and NMT developers, a key area for improvement is cultural and contextual adaptation. AI models should be trained on a larger corpus of idiomatic expressions, including proverbs, metaphors, and culturally specific phrases. DeepSeek’s translation of “*You scratch my back and I scratch yours*” as “*المصالح متبادلة*” (Mutual interests) demonstrated its ability to choose culturally relevant alternatives, while ChatGPT’s literal translation lacked idiomatic nuance. Additionally, AI systems must improve their ability to detect and preserve metaphors, as seen in the case of “*Passing the exam is not a bed of roses.*” DeepSeek captured the metaphor with “*ليس كالسير على الورود*” (Not like walking on roses), while ChatGPT’s translation was more literal and missed the figurative meaning.

Another recommendation is to optimize for natural Arabic fluency. NMT models should prioritize idiomatic and culturally resonant phrasing, as demonstrated by DeepSeek’s “*لنقلب صفحة جديدة*” (Let’s turn a new page), which sounds more natural than ChatGPT’s more formal translation. (See table 28 in appendix E).

Addressing identified weaknesses is essential as well. Many idioms lack direct equivalents in the target language, as seen in the translation of “*They signed on the dotted line.*” DeepSeek’s literal “*وقعوا على الخط المنقط*” (They signed on the dotted line) was more accurate than ChatGPT’s misrepresentation. AI models should leverage functional equivalents where direct translations do not exist, while ensuring consistency in formality or register, as demonstrated by DeepSeek’s more formal translation of “*Lend me your ears*” as “*أصغوا إليّ*” (Lend me your ear), which did not match the original idiom’s tone.

For translators and educators, AI can be a powerful collaborative tool. Accordingly, translators should consider combining outputs from both ChatGPT and DeepSeek to select the best translation for each idiom. For instance, in cases where a phrase like “*They signed on the dotted line*” lacks idiomatic resonance, DeepSeek’s more literal approach might work better, while ChatGPT’s functional translation could be enhanced with post-editing to add cultural nuance. Accordingly, post-editing should focus on idioms that lack direct equivalents or fail to capture the cultural meaning. The aim is to ensure that the final translation is both idiomatic and accurate. Educators should incorporate these AI-driven insights into their curriculum by teaching translation strategies, such as ‘borrowing’ for idioms with direct equivalents and ‘adaptation’ for culture-specific expressions. Using categorized examples from this study, educators can demonstrate ideal translations and

offer students guidance on adapting idioms that lack direct cultural equivalents.

For future research, expanding the dataset to include more idiomatic expressions from various categories as well as various Arabic dialects and specialized domains would provide valuable insights into how NMT systems handle idioms in different contexts. Additionally, testing newer NMT models, such as Gemini or localized Arabic large language models (LLMs), could improve idiomatic translation capabilities and cultural adaptation. Human evaluation is also crucial, as native Arabic speakers can assess the naturalness and cultural appropriateness of machine-generated translations. This offers valuable feedback to refine these systems.

In conclusion, improving machine translation for idiomatic expressions requires attention to various points. These are: cultural and contextual nuances, natural fluency, and precise adaptation of idioms. Furthermore, the effectiveness of machine translation can be significantly enhanced via various processes. These are; (1) combining AI outputs with human expertise, (2) developing robust training programs, and (3) expanding research into diverse dialects and domains. To sum up, the field of machine translation remains a fertile soil, yet to be ploughed with various methods and processes for improvements.

Acknowledgments

My sincere thanks and gratitude are to Almighty Allah for enabling me to successfully conduct this study.

Author contributions

Dr. Shaimaa Mohamed Helal was responsible for the whole paper.

Funding

The author extends her appreciation to the Arab Open University for funding this work through AOU research fund No. (AOUKSA-524008).

Competing interests

“Not applicable.”

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal’s policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Aalfaqih, A. (2025). THE EFFECTIVENESS OF AI TOOLS IN EVALUATING IDIOMATIC EXPRESSIONS USAGE AMONG LEARNERS. *GLOBAL: Education Language and Humanity Journal*, 2(2), 66-85. <https://doi.org/10.66355/95gfw07>
- Ahmad, M., & Khan, S. (2020). Dynamic equivalence in English-Urdu poetry translation: A comparative study. *Journal of Translation Studies*, 15(3), 78-95.
- AlAfnan, M. A. (2025). DeepSeek Vs. ChatGPT: A Comparative Evaluation of AI Tools in Composition, Business Writing, and Communication Tasks. *Journal of Artificial Intelligence and Technology*. <https://doi.org/10.37965/jait.2025.0740>
- Albogdady, A. (2023). Translating Shakespeare’s Midsummer Night’s Dream’s Latin Rhetorical Schemes in the light of Vinay and Darbelnet’s Translation Theory. *Mağallāī Kullīyyāī Al-Adāb - Ğāmi‘at Al-Fayūm*, 15(2), 210-271. <https://doi.org/10.21608/jfafu.2023.206009.1898>
- Al-Darraji, O. A. O. O. (2015). Equivalence in translation: An investigation of the usefulness of formal and dynamic types of equivalence in

- the translation of some English idiomatic expressions into Arabic. *Global Libyan Journal*, 12(2), 234-251.
- Al-Fdhl, S. (2023). Vinay and darbelnet's translating procedures used in translating culture-specific words: A case study of george orwell's animal farm. *Electronic Journal of University of Aden for Humanity and Social Sciences*. <https://doi.org/10.47372/ejua-hs.2023.1.235>
- Almahasees, Z. (2021). *Analysing English-Arabic Machine Translation: Google Translate, Microsoft Translator and Sakhr*. Routledge. <https://doi.org/10.4324/9781003191018>
- Al-Mizgagi, E. H., & Al-Tamemi, J. A. (2023). *Correlation between strategies of translation and politeness in Arms and the Man*. <https://doi.org/10.59222/ustjmhs.1.4.4>
- Alshammari, J. N. M. (2016). Examining Nida's translation theory in rendering Arabic proverbs into English: A comparative analysis study. *Journal of Applied Linguistics and Language Research*, 3(1), 156-172.
- Anderson, P., Thompson, K., & Wilson, R. (2021). Optimizing translation memory systems for dynamic equivalence applications. *Translation Technology Review*, 8(2), 45-62.
- Aziz, M. (2022). Applying Vinay and Darbelnet's translational procedures in dubbing animation from English into Arabic. *English Linguistics Research*, 11(2). <https://doi.org/10.5430/elr.v11n2p22>
- Baker, M. (2018). *In other words: A coursebook on translation*. Routledge. <https://doi.org/10.4324/9781315619187>
- Bassnett, S. (2013). *Translation studies*. Routledge. <https://doi.org/10.4324/9780203488232>
- Baziotis, C., Mathur, P., & Hasler, E. (2023). Automatic evaluation and analysis of idioms in neural machine translation. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, 3682–3700. <https://doi.org/10.18653/v1/2023.eacl-main.267>
- Cao, P., Khoshsaligheh, M., & Jomhour, F. (2025). Deepseek, ChatGPT, and Gemini versus human subtitling: a case study in socio-cultural adaptation in multimedia communication. *Perspectives*, 1-19. <https://doi.org/10.1080/0907676X.2025.2562024>
- Chan, V., & Tang, K. W. W. (2024a). GPT for translation: a systematic literature review. *SN Comput. Sci.*, 5, 986. <https://doi.org/10.1007/s42979-024-03340-z>
- Chen, W., & Nakamura, H. (2020). Buddhist sutra translation: Dynamic equivalence and spiritual meaning preservation. *Asian Translation Studies*, 7(1), 23-41.
- Choi, W. C., & Chang, C. (2025). Advantages and Limitations of Open-Source Versus Commercial Large Language Models (LLMs): A Comparative Study of DeepSeek and OpenAI's ChatGPT. <https://doi.org/10.20944/preprints202503.1081.v1>
- Dwipusita, W., & Wediyantoro, P. L. (2023). Translation strategies in Shakespeare's Romeo and Juliet translated drama script viewed from the romance perspective. *EnJourMe (English Journal of Merdeka): Culture, Language, and Teaching of English*, 8(2). <https://doi.org/10.26905/enjourme.v8i2.10950>
- El-Farhaty, H. (2015). *Arabic-English-Arabic Legal Translation*. New York: Routledge. <https://doi.org/10.4324/9781315745893>
- 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. <https://doi.org/10.1080/23311886.2024.2410998>
- Ghazala, H. S. (2008). *Translation as problems and solutions: A textbook for university students and trainee translators*. Special Edition. DAR ELILM LILMALAYN.
- Guo, L., & Amirdabbaghian, A. (2025). Exploring the Translation of Idioms by AI in Chinese Internet Literature. *In Role of AI in Translation and Interpretation* (pp. 293-326). IGI Global Scientific Publishing. <https://doi.org/10.4018/979-8-3373-0060-3.ch011>
- Haider, A. S., Darwish, N., Saed, H., Dagamseh, M., Alantari, D., & Tannous, B. (2025). A reception study of AI-translated idioms and proverbs between Arabic and English. *Research Journal in Advanced Humanities*, 6(3). <https://doi.org/10.58256/k4d6pp20>
- Hajjo, R., Sabbah, D. A., & Bardaweel, S. K. (2025). *ChatGPT vs. DeepSeek: A Comparative Analysis of AI Models for Breast Cancer Information Retrieval* (Preprint). <https://doi.org/10.2196/preprints.72839>
- Hani, H. A., & Zainab, T. (2024). The application of Vinay and Darbelnet model to English and Urdu translations of Surah Ad-Duha: A comparative study. *Journal of Applied Linguistics*, 3(2). <https://doi.org/10.52622/joal.v3i2.176>
- Hassan, G. B. M., & Hussein, B. A. (2024). The Mechanisms of Literary Translation in the Translation of "امشب اشكى ميريزد" from Persian to Kurdish according to the Theories of (Vinay) and (Darbelnet). *II*. *كۆنفرانسى زانك فۆى رابھه درين*, 6(6), 507-531. [https://doi.org/10.26750/vol\(11\).no\(6\).paper20](https://doi.org/10.26750/vol(11).no(6).paper20)
- Hatim, B., & Munday, J. (2019). *Translation: An advanced resource book for students*. Routledge. <https://doi.org/10.4324/9780429266348>
- Huang, J., & Wang, G. (2024). A study on translation strategies of legal English texts from the perspective of functional equivalence theory. *Frontiers in Humanities and Social Sciences*, 12(3), 156-171. <https://doi.org/10.54691/pycvvc29>
- Jabak, O. (2020). Application of Eugene Nida's theory of translation to the English translation of Surah Ash-Shams. *TranscUlturAl: A Journal of Translation and Cultural Studies*, 12(2), 89-107. <https://doi.org/10.21992/tc29461>

- Jiang, Y. (2020). Critical thinking of applying Nida's functional equivalence to poetry translation: A study based on English versions of Lu Zhai by Wang Wei. *Studies in Literature and Language*, 21(2), 45-58.
- Jibreel, I. (2023). Online Machine Translation Efficiency in Translating Fixed Expressions Between English and Arabic (Proverbs as a Case-in-Point). *Theory and Practice in Language Studies*, 13(5), 1148-1158. <https://doi.org/10.17507/tpls.1305.07>
- Kocabiyik, H. (2022). The analysis of the Turkish translation of the figurative language in the novel "The Day The Leader Was Killed" by Nejb Mahfouz within the ambit of theories coined by Vinay and Darbelnet, Newmark, Pym and Toury. *RumeliDE Dil ve Edebiyat Arařtırmaları Dergisi*. <https://doi.org/10.29000/rumelide.1132601>
- Kotsis, K. T. (2025). Comparing ChatGPT and DeepSeek in Addressing Misconceptions about Physics Concepts. *European Journal of Contemporary Education and E-Learning*, 3(2), 191-206. [https://doi.org/10.59324/ejceel.2025.3\(2\).17](https://doi.org/10.59324/ejceel.2025.3(2).17)
- Li, X., & Zhang, Y. (2021). Medical text translation: Dynamic equivalence and patient comprehension. *Medical Translation Journal*, 18(4), 234-251.
- Liu, D. (2017). *Idioms: Description, comprehension, acquisition, and pedagogy*. Routledge. <https://doi.org/10.4324/9781315092843>
- Liu, E., Chaudhary, A., & Neubig, G. (2023). Crossing the threshold: Idiomatic machine translation through retrieval augmentation and loss weighting. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, 15095-15111. <https://doi.org/10.18653/v1/2023.emnlp-main.933>
- Mahyoob, M., & Algaraady, J. (2025). *DeepSeek vs. ChatGPT: Comparative Efficacy in Reasoning for Adults' Second Language Acquisition Analysis*. https://doi.org/10.31235/osf.io/fapmc_v1
- Martinez, C., Rodriguez, A., & Silva, M. (2022). Dynamic equivalence in Spanish-English medical document translation. *Healthcare Translation Review*, 9(3), 145-162.
- Mounassar, A. A. A. (2024). The effectiveness of machine translation systems in translating English fixed phrases into Arabic. *University of Science and Technology Journal for Management and Human Sciences*, 2(2), 91-116. <https://doi.org/10.59222/ustjmhs.2.2.4>
- Munday, J. (2016). *Introducing Translation Studies: Theories and Applications* (4th ed.). London and New York: Routledge. <https://doi.org/10.4324/9781315691862>
- Nemati, A., & Afzalifard, E. (2017). An investigation into Farsi translation of children's poetry of Shel Silverstein according to Nida's theory. *Journal of Applied Linguistics and Language Research*, 4(1), 112-128.
- Nida, E. A. (1964). *Toward a science of translating: with special reference to principles and procedures involved in Bible translating*. Brill Archive. <https://doi.org/10.1163/9789004495746>
- Nida, E. A. (2019). Language and culture in Eugene Nida's work the dynamic equivalence: Critics and defenders. *Journal of Modern Education Review*, 9(5), 234-248. [https://doi.org/10.15341/jmer\(2155-7993\)/05.09.2019/004](https://doi.org/10.15341/jmer(2155-7993)/05.09.2019/004)
- Nida, E. A., & Taber, C. R. (1969). *The theory and practice of translation*. Brill.
- Nunberg, G., Sag, I. A., & Wasow, T. (1994). Idioms. In S. Everson (Ed.), *Language* (pp. 491-538). Cambridge University Press. <https://doi.org/10.2307/416483>
- Pietryga, M. (2020). Application of formal and dynamic equivalence to chosen press releases published by the European Commission related to the issue of the coronavirus pandemic. *Tertium*, 5(1), 67-84. <https://doi.org/10.7592/Tertium2020.5.1.Pietryga>
- Pym, A. (2023). *Exploring translation theories*. Routledge. <https://doi.org/10.4324/9781003383130>
- Rahman, A. (2021). Hadith translation and dynamic equivalence: Bridging religious concepts across cultures. *Islamic Translation Studies*, 6(2), 45-63.
- Rodriguez, M., & Schmidt, K. (2021). Cross-linguistic patterns in dynamic equivalence: A comparative corpus study. *Corpus Translation Studies*, 15(3), 112-129.
- Safi, S. H., & Nasser, L. A. (2022). Introduction to The Model of Vinay and Darbelnet in Translation: Basics and Comparisons. *College of Basic Education Researches Journal*, 18(3). <https://doi.org/10.33899/berj.2022.175705>
- Sari, D., Putri, L., & Wahyuni, S. (2019). English-Indonesian subtitle translation: Dynamic equivalence and cultural authenticity. *Audiovisual Translation Journal*, 11(2), 67-84.
- Shuttleworth, M., & Cowie, M. (2014). *Dictionary of Translation Studies*. St Jerome <https://doi.org/10.4324/9781315760490>
- Sickinger, P. (2017). Aiming for cognitive equivalence: Mental models as a tertium comparationis for translation and empirical semantics. *Research in Language*, 15(2), 123-141. <https://doi.org/10.1515/rela-2017-0013>
- Siefring, J. (2004). *Oxford Dictionary of Idioms* (2nd ed.). Oxford: Oxford University Press.
- Silva, R. (2019). Medieval English-Portuguese translation: Dynamic equivalence across temporal distance. *Historical Translation Studies*, 12(1), 34-52.

- Sufen, L. (2019). Cultural differences and translation strategies in business English translation under the guidance of functional equivalence principle. *International Language Education and Literature Conference*, 456-471.
- Sun, Y., & Wan, P. (2022). News translation from the perspective of functional equivalence theory in China Daily. *Study in English Language Teaching*, 10(4), 89-104. <https://doi.org/10.22158/selt.v10n4p82>
- Suprayetno, E., Ajmal, M. U., Khan, M., Al, S., & Langkat - Indonesia, M. M. (2024). A comparative analysis of "Jawab e Shikwa": A Vinay and Darbelnet model-based examination of two English translations. *Journal of Applied Linguistics*, 3(2). <https://doi.org/10.52622/joal.v3i2.171>
- Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open*, 1, 5-21. <https://doi.org/10.1016/j.aiopen.2020.11.001>
- Tavakoli, S., & Zare, M. (2024). Bridging Cultures through Verse: Shakespearean Sonnets' Persian Translations Through the Lens of Vinay-Darbelnet's Model. *İstanbul Üniversitesi Çeviribilim Dergisi*, 0(20), 122-137. <https://doi.org/10.26650/ijts.2024.1452977>
- Taylor, S., & Johnson, M. (2020). Cognitive processing in dynamic equivalence translation: An eye-tracking study. *Cognitive Translation Studies*, 7(2), 78-95.
- Thompson, R., & Williams, D. (2018). Protestant Bible translation and reader comprehension: A dynamic equivalence analysis. *Biblical Translation Studies*, 22(3), 145-162.
- Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.
- Vinay, J. P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation* (J. C. Sager & M. J. Hamel, Trans.). John Benjamins. (Original work published 1958). <https://doi.org/10.1075/btl.11>
- Vyas, H., & Bhardwaj, R. G. (2025). *ChatGPT vs DeepSeek: A Comparative Evaluation on the International Computer Science Benchmark - ACM ICPC*. <https://doi.org/10.21203/rs.3.rs-7077588/v1>
- Wang, L., & Li, H. (2023). Machine translation post-editing through dynamic equivalence frameworks. *Computational Translation Studies*, 16(4), 234-251.
- Zaitova, I., et al. (2025). It's not a walk in the park! *Challenges of idiom translation in speech-to-text systems*. arXiv. <https://doi.org/10.18653/v1/2025.acl-long.1512>
- Zhang, R., Cai, Q., Sartori, A. L. B., Gayed, N., & Collette, H. (2025). *Comparing Artificial Intelligence Large Language Models in Medical Training: A Performance Analysis of ChatGPT and DeepSeek on United States Medical Licensing Examination (USMLE) Style Questions*. Cureus. <https://doi.org/10.7759/cureus.90212>
- Ziyi, W., & Zheng, C. (2023). A report on E-C translation of The Lord of the Rings based on functional equivalence theory. *US-China Foreign Language*, 21(11), 156-173. <https://doi.org/10.17265/1539-8080/2023.11.006>

Appendices

Appendix A: Data

Translate from English into Arabic:

- 1- Passing the exam is not a bed of roses
- 2- A true friend does not stab in the back
- 3- You scratch my back and I scratch yours
- 4- He killed his neighbour in cold blood
- 5- You make my blood boil by your bad manners
- 6- It was the straw that broke the camel's back
- 7- High prices are a daylight robbery
- 8- Our dear aunt is at death's door
- 9- They needed the job, so they signed on the dotted line
- 10- She cannot believe her eyes/ears
- 11- Their company is on the black list
- 12- Why do you wash your dirty linen in public
- 13- Some people put the cart before the horse
- 14- That man is rolling in money
- 15- People were driven to war like lambs to the slaughter
- 16- They apply the law of the jungle
- 17- Ladies and gentlemen, lend me your ears
- 18- He serves two masters
- 19- Let us turn a new page
- 20- His name was on the tip of my tongue

Appendix B: Screenshots of DeepSeek & ChatGPT

Figure 1. Screenshot of DeepSeek (1)

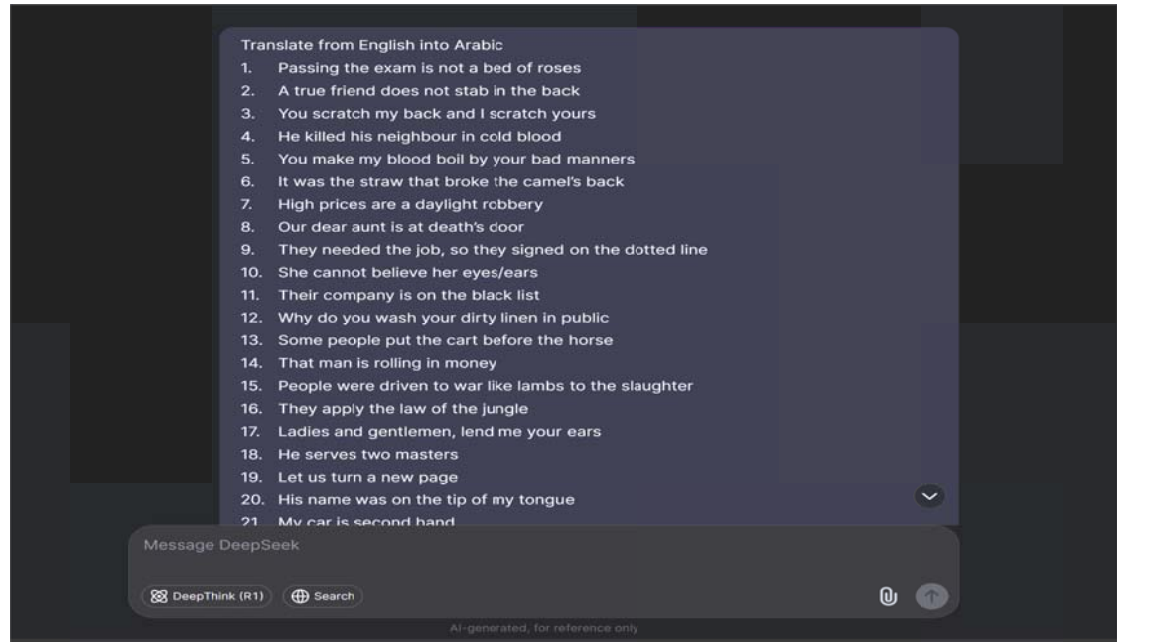


Figure 2. Screenshot of DeepSeek (2)



Figure 3. Screenshot of ChatGPT (1)

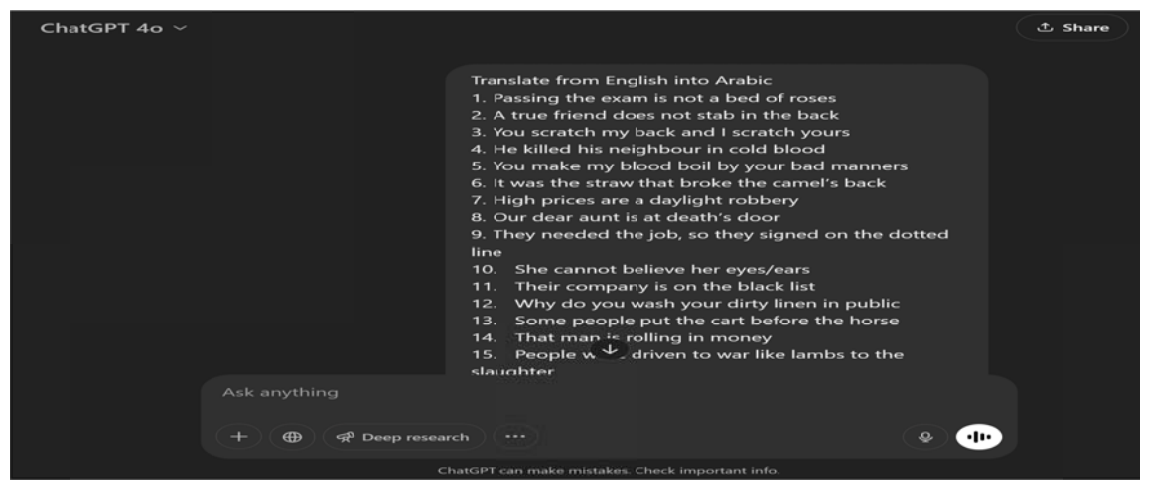


Figure 4. Screenshot of ChatGPT (2)

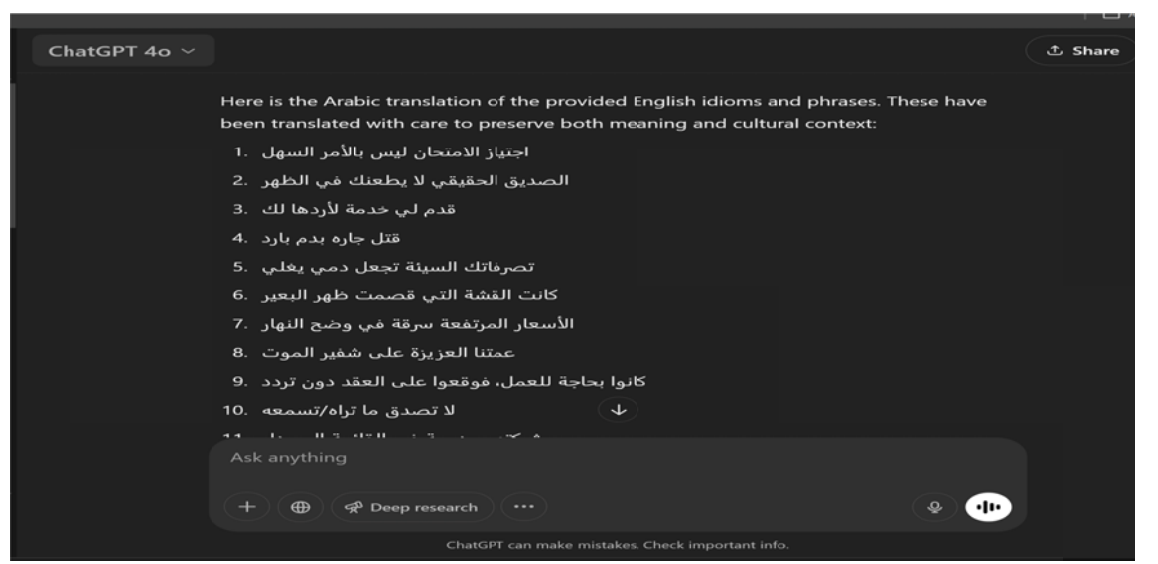
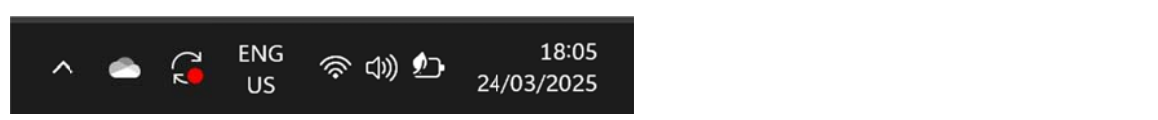


Figure 5. Screenshot of Date of submission to DeepSeek & ChatGPT (24-3-2025)



Appendix C: Quantitative Evaluation

Table 13. Quantitative Evaluation of Translations Generated by ChatGPT and DeepSeek

No.	English Sentence/idiom	ChatGPT Translation	DeepSeek Translation	Criteria	ChatGPT Analysis & Points	DeepSeek Analysis & Points
1	Passing the exam is not a bed of roses	اجتياز الامتحان ليس بالأمر السهل	اجتياز الامتحان ليس سهلاً (أو): ليس كالسير على الورود	Accuracy	Accurate but loses metaphor. (0 pts)	Offers accurate literal and metaphorical option. (1 pt)
				Cult. Appropriateness	Lacks cultural equivalence. (0 pts)	”ليس كالسير على الورود“ is culturally suitable. (1 pt)
				Naturalness	Readable but flat. (0 pts)	More natural due to idiomatic option. (1 pt)
2	A true friend does not stab in the back	الصديق الحقيقي لا يطعنك في الظهر	الصديق الحقيقي لا يطعنك في الظهر	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
3	You scratch my back and I scratch yours	قدم لي خدمة لأردّها لك	أنت تساعدني وأساعدك (أو): المصالح متبادلة	Accuracy	Accurate but lacks idiom. (0 pts)	Better captures idiom with ”المصالح متبادلة“. (1 pt)
				Cult. Appropriateness	Functional but not idiomatic. (0 pts)	Culturally appropriate phrase. (1 pt)
				Naturalness	Readable but flat. (0 pts)	Natural and idiomatic. (1 pt)
4	He killed his neighbour in cold blood	قتل جاره بدم بارد	قتل جاره بدم بارد	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)

No.	English Sentence/idiom	ChatGPT Translation	DeepSeek Translation	Criteria	ChatGPT Analysis & Points	DeepSeek Analysis & Points
5	You make my blood boil...	تصرفاتك السيئة تجعل دمي يغلي	تصرفاتك السيئة تغلي دمي	Accuracy	Accurate and idiomatic. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Acceptable but less common. (0.5 pts)
				Naturalness	Natural and fluent. (1pt)	Slightly awkward phrasing. (0 pts)
6	the straw that broke the camel's back	كانت القشة التي قصمت ظهر البعير	كانت تلك القشة التي قصمت ظهر البعير	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
7	High prices are a daylight robbery	الأسعار المرتفعة سرقة في وضح النهار	الأسعار المرتفعة هي سرقة في وضح النهار	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
8	Our dear aunt is at death's door	عمتنا العزيزة على شفير الموت	عمتنا العزيزة على أعتاب الموت	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
9	They signed on the dotted line	وقعوا على العقد دون تردد	وقعوا على الخط المنقط	Accuracy	Misrepresents the idiom. (0 pts)	More accurate but literal. (1 pt)
				Cult. Appropriateness	Lacks cultural equivalence. (0 pts)	Functional but not idiomatic. (0 pts)

No.	English Sentence/idiom	ChatGPT Translation	DeepSeek Translation	Criteria	ChatGPT Analysis & Points	DeepSeek Analysis & Points
				Naturalness	Natural but less precise. (0 pts)	Accurate but lacks fluency. (0 pts)
10	She cannot believe her eyes/ears	لا تصدق ما تراه/تسمعه	لا تصدق ما تراه عيناها/تسمعه أذناها	Accuracy	Accurate and concise. (1 pt)	Accurate but overly formal. (0 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	More natural. (1 pt)	Slightly less natural. (0 pts)
11	Their company is on the black list	شركتهم مدرجة في القائمة السوداء	شركتهم على القائمة السوداء	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
12	Why do you wash your dirty linen in public?	لماذا تغسل غسيلك القذر أمام الناس؟	لماذا تنشر مشاكلك أمام الجميع؟	Accuracy	Accurate but literal. (0 pts)	Adapts idiomatically. (1 pt)
				Cult. Appropriateness	Slightly off culturally. (0 pts)	More culturally appropriate. (1 pt)
				Naturalness	Literal but less fluent. (0 pts)	Natural and idiomatic. (1 pt)
13	Some people put the cart before the horse	بعض الناس يضعون العربة أمام الحصان	بعض الناس يضعون العربة أمام الحصان	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
14	That man is rolling in money	ذلك الرجل يفرق في المال	ذلك الرجل يسبح في الأموال	Accuracy	Accurate but less idiomatic. (0 pts)	More relevant idiomatic phrase. (1 pt)

No.	English Sentence/idiom	ChatGPT Translation	DeepSeek Translation	Criteria	ChatGPT Analysis & Points	DeepSeek Analysis & Points
				Cult. Appropriateness	Acceptable but less common. (0 pts)	More culturally appropriate. (1 pt)
				Naturalness	Slightly flat. (0 pts)	Natural and idiomatic. (1 pt)
15	People were driven to war like lambs...	سُيِّقَ الناس إلى الحرب كأنهم خراف إلى الذبح	تم دفع الناس إلى الحرب كالأغنام إلى الذبح	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
16	They apply the law of the jungle	يطبقون قانون الغاب	يُطَبِّقون قانون الغاب	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
17	Ladies and gentlemen, lend me your ears	سيدياتي وساداتي، استمعوا إلي	سيدياتي ساداتي، أصغوا إلي	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate, more formal. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
18	He serves two masters	يخدم سيدين في آن واحد	هو يخدم سيدين في آن واحد	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)

No.	English Sentence/idiom	ChatGPT Translation	DeepSeek Translation	Criteria	ChatGPT Analysis & Points	DeepSeek Analysis & Points
19	Let us turn a new page	دعونا نفتح صفحة جديدة	لنقلب صفحة جديدة	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)
20	His name was on the tip of my tongue	كان اسمه على طرف لساني	كان اسمه على طرف لساني	Accuracy	Accurate. (0.5 pts)	Accurate. (0.5 pts)
				Cult. Appropriateness	Culturally appropriate. (0.5 pts)	Culturally appropriate. (0.5 pts)
				Naturalness	Natural. (0.5 pts)	Natural. (0.5 pts)

Appendix D: Qualitative Analysis Based on Three Criteria

Table 14. Translations Generated by ChatGPT and DeepSeek

No.	English Sentence	ChatGPT Translation	Deep Seek Translation	Accuracy	Cultural Appropriateness	Naturalness
1	Passing the exam is not a bed of roses	اجتياز الامتحان ليس بالأمر السهل	اجتياز الامتحان ليس سهلاً (أو: ليس كالسير على الورود)	ChatGPT: Accurate but does not convey the metaphorical meaning of "bed of roses." Deep Seek: Offers a more idiomatic alternative with "كالسير على الورود."	ChatGPT: Lacks cultural equivalence; literal translation. Deep Seek: "كالسير على الورود" is culturally suitable.	ChatGPT: Readable but somewhat flat. Deep Seek: More natural due to the idiomatic addition.
2	A true friend does not stab in the back	الصديق الحقيقي لا يطعنك في الظهر	الصديق الحقيقي لا يطعنك في الظهر	Both translations are accurate, conveying the intended meaning directly.	Both translations are culturally appropriate as "طعنك في الظهر" aligns with Arabic idiomatic norms.	Both translations are fluent and natural.
3	You scratch my back and I scratch yours	قدم لي خدمة لأردھا لك	أنت تساعدني وأساعدك (أو: المصالح متبادلة)	ChatGPT: Accurate but lacks idiomatic flavor. Deep Seek: Adds "المصالح متبادلة," which better captures the idiom.	ChatGPT: Functional but not culturally idiomatic. Deep Seek: Culturally appropriate phrase, "المصالح متبادلة."	ChatGPT: Readable but flat. Deep Seek: Natural and idiomatic.
4	He killed his neighbour in cold blood	قتل جاره بدم بارد	قتل جاره بدم بارد	Both translations are accurate and convey the intended meaning.	Both translations align culturally as "بدم بارد" is a common Arabic idiom.	Both translations are fluent and idiomatic.
5	You make my blood boil by your bad manners	تصرفاتك السيئة تجعل دمي يغلي	تصرفاتك السيئة تغلي دمي	ChatGPT: Accurate and idiomatic. Deep Seek: Slightly less natural due to word order "تغلي دمي."	ChatGPT: Culturally appropriate. Deep Seek: Acceptable but less common word order.	ChatGPT: Natural and fluent. Deep Seek: Slightly awkward phrasing.
6	It was the straw that broke the camel's back	كانت القشة التي قصمت ظهر البعير	كانت تلك القشة التي قصمت ظهر البعير	Both translations are accurate.	Both translations are culturally appropriate as the idiom exists in Arabic in the same form.	Both translations are idiomatic and natural.

No.	English Sentence	ChatGPT Translation	Deep Seek Translation	Accuracy	Cultural Appropriateness	Naturalness
7	High prices are a daylight robbery	الأسعار المرتفعة سرقة في وضح النهار	الأسعار المرتفعة هي سرقة في وضح النهار	Both translations are accurate, though Deep Seek's phrasing is slightly more formal.	Both are culturally appropriate as "سرقة في وضح النهار" aligns with Arabic idiomatic usage.	Both translations are natural and fluent.
8	Our dear aunt is at death's door	عمتنا العزيزة على شفير الموت	عمتنا العزيزة على أعتاب الموت	Both translations are accurate, with slight variation in phrasing.	Both translations are culturally appropriate. "على شفير الموت" and "على أعتاب الموت" are interchangeable in Arabic.	Both translations are natural and readable.
9	They signed on the dotted line	وقعوا على العقد دون تردد	وقعوا على الخط المنقط	ChatGPT: Misrepresents the idiom. Deep Seek: More accurate but literal.	ChatGPT: Lacks cultural equivalence. Deep Seek: Functional but not idiomatic.	ChatGPT: Natural but less precise. Deep Seek: Accurate but lacks idiomatic fluency.
10	She cannot believe her eyes/ears	لا تصدق ما تراه/تسمعه	لا تصدق ما تراه عيناها/تسمعه أذناها	ChatGPT: Accurate and concise. Deep Seek: Accurate but overly formal.	Both translations are culturally appropriate.	ChatGPT: More natural. Deep Seek: Slightly less natural due to redundancy.
11	Their company is on the black list	شركتهم مدرجة في القائمة السوداء	شركتهم على القائمة السوداء	Both translations are accurate and convey the intended meaning.	Both translations are culturally appropriate as "القائمة السوداء" is widely used in Arabic.	Both translations are fluent and natural.
12	Why do you wash your dirty linen in public	لماذا تغسل غسيلك القذر أمام الناس؟	لماذا تنتشر مشاكلك أمام الجميع؟	ChatGPT: Accurate but literal. Deep Seek: Adapts idiomatically with "تنتشر مشاكلك".	ChatGPT: Slightly off culturally. Deep Seek: More culturally appropriate.	ChatGPT: Literal but less fluent. Deep Seek: Natural and idiomatic.
13	Some people put the cart before the horse	بعض الناس يضعون العربة أمام الحصان	بعض الناس يضعون العربة أمام الحصان	Both translations are accurate and convey the intended meaning.	Both are culturally appropriate as the idiom is used in Arabic in the same way.	Both translations are natural and fluent.
14	That man is rolling in money	ذلك الرجل يغرق في المال	ذلك الرجل يسبح في الأموال	ChatGPT: Accurate but less idiomatic. Deep Seek: Offers a more relevant idiomatic phrase.	ChatGPT: Acceptable but less common. Deep Seek: More culturally appropriate.	ChatGPT: Slightly flat. Deep Seek: Natural and idiomatic.
15	People were driven to war like lambs to the slaughter	سُيق الناس إلى الحرب كأنهم خراف إلى الذبح	تم دفع الناس إلى الحرب كالأغنام إلى الذبح	Both translations are accurate.	Both translations are culturally appropriate, though "خراف" is more common than "الأغنام" in idiomatic use.	Both translations are fluent and idiomatic.
16	They apply the law of the jungle	يطبقون قانون الغاب	يُطبِقون قانون الغاب	Both translations are accurate and convey the intended meaning.	Both translations are culturally appropriate.	Both translations are fluent and idiomatic.
17	Ladies and gentlemen, lend me your ears	سيداتي وسادتي، استمعوا إلي	سيداتي سادتي، أصغوا إلي	Both translations are accurate.	Both translations are culturally appropriate, though "أصغوا" is slightly more formal.	Both translations are fluent and natural.
18	He serves two masters	يخدم سيدين في آن واحد	هو يخدم سيدين في آن واحد	Both translations are accurate, with slight differences in phrasing.	Both translations are culturally appropriate.	Both translations are fluent and natural.
19	Let us turn a new page	دعونا نفتح صفحة جديدة	لنقلب صفحة جديدة	Both translations are accurate.	Both translations are culturally appropriate.	Both translations are fluent and natural.
20	His name was on the tip of my tongue	كان اسمه على طرف لساني	كان اسمه على طرف لساني	Both translations are accurate and identical.	Both translations are culturally appropriate.	Both translations are fluent and natural.

Appendix E: Qualitative Analysis based on Nida’s Dynamic Equivalence, Vinay & Darbelnet’s Strategies, and NMT Performance.

Table 15. Example 2: “A true friend does not stab in the back.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations (لا يطعنك في الظهر) achieve perfect dynamic equivalence. The idiom is a direct calque that exists identically in Arabic, ensuring the meaning and the negative connotation of betrayal are perfectly preserved for the target audience.
Vinay & Darbelnet’s Strategies	Both models correctly apply a Calque (a literal translation of the idiom that is already assimilated into the target language). This is the optimal strategy here.
NMT Performance	Both systems perform flawlessly, recognizing a perfect cross-linguistic equivalent.

Table 16. Example 4: “He killed his neighbour in cold blood.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations (بدم بارد) achieve full dynamic equivalence. The calque is completely naturalized in Arabic and carries the exact same meaning of a ruthless, premeditated act.
Vinay & Darbelnet’s Strategies	Both models use a Calque.
NMT Performance	Perfect performance from both systems.

Table 17. Example 5: “You make my blood boil by your bad manners.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations are literally accurate and understood. The metaphor is common in Arabic as this is direct metaphor (Gazala, 2008). It means تغضبني بشدة (you make me very angry). Therefore, equivalence is achieved.
Vinay & Darbelnet’s Strategies	Both models use Literal Translation.
NMT Performance	Both systems default to a literal translation which here is acceptable as the same metaphor is applicable in Arabic. انك تجعل الدم يغلي في عروقي بسلوكك السيئ as given by Gazala (2008, p. 128).

Table 18. Example 6: “It was the straw that broke the camel’s back.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations (القشة التي قصمت ظهر البعير) achieve perfect dynamic equivalence. The idiom is identical in English and Arabic, making the translation seamless and naturally evocative for the target audience.
Vinay & Darbelnet’s Strategies	Both models use a Calque.

Theory	Analysis
NMT Performance	Excellent performance from both. DeepSeek’s addition of تلك(that) adds a slight emphatic nuance, mirroring the definiteness of the English “the.”

Table 19. Example 7: “High prices are a daylight robbery.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations (سرقة في وضح النهار) achieve dynamic equivalence. The calque is perfectly naturalized in Arabic and conveys the same sense of blatant exploitation.
Vinay & Darbelnet’s Strategies	Both models use a Calque.
NMT Performance	Both systems perform correctly.

Table 20. Example 8: “Our dear aunt is at death’s door.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations achieve dynamic equivalence. على شفير الموت (on the brink of death) and على أعتاب الموت (on the threshold of death) are synonymous, culturally appropriate metaphors that convey the same imminent sense of death.
Vinay & Darbelnet’s Strategies	Both models use Modulation, substituting the original metaphor (“door”) with culturally synonymous images (“brink,” “threshold”).
NMT Performance	Both systems perform well, demonstrating an understanding of nuanced metaphorical equivalence.

Table 21. Example 10: “She cannot believe her eyes/ears.”

Theory	Analysis
Nida’s Dynamic Equivalence	ChatGPT’s translation (لا تصدق ما تراه/تسمعه) achieves better dynamic equivalence as it reflects how the concept is naturally expressed in Arabic. DeepSeek’s more literal version (لا تصدق ما تراه عيناها) is accurate but slightly redundant and formal.
Vinay & Darbelnet’s Strategies	ChatGPT uses Modulation, simplifying the structure to its natural form. DeepSeek uses Literal Translation.
NMT Performance	ChatGPT demonstrates a better grasp of natural Arabic phrasing in this instance.

Table 22. Example 11: “Their company is on the black list.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations (على/في القائمة السوداء) achieve full dynamic equivalence. The term is a direct, fully assimilated calque in Arabic. The same translation is given by Ghazala (2008) : “شركتهم على القائمة السوداء” (p. 129).
Vinay & Darbelnet’s	Both models use a Calque.

Theory	Analysis
Strategies	
NMT Performance	Perfect performance.

Table 23. Example 13: “Some people put the cart before the horse.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations (يضعون العربيه أمام الحصان) achieve perfect dynamic equivalence. The idiom is an established calque in Arabic. The same translation is given by Ghazala (2008): “يضع بعض الناس العربيه امام الحصان” (p. 129).
Vinay & Darbelnet’s Strategies	Both models use a Calque.
NMT Performance	Perfect performance.

Table 24. Example 15: “People were driven to war like lambs to the slaughter.”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations achieve dynamic equivalence. خراف(lambs) and الأغنام(sheep) are both used in Arabic to convey innocence and helplessness in this context. The metaphors are equally powerful.
Vinay & Darbelnet’s Strategies	Both models use Literal Translation of an established metaphorical comparison.
NMT Performance	Both systems perform correctly.

Table 25. Example 16: “They apply the law of the jungle”

Theory	Analysis
Nida’s Dynamic Equivalence	Both translations achieve perfect dynamic equivalence. The idiom “law of the jungle” has a direct, culturally understood equivalent in Arabic (قانون الغاب). The translations successfully convey the concept of ruthless, survival-of-the-fittest behavior, evoking the same response in an Arabic reader as the original does in an English reader.
Vinay & Darbelnet’s Strategies	This is a clear example of Equivalence. The translators did not translate the words “jungle” or “law” literally (قانون الغابة would be more literal but less idiomatic). Instead, they used the standard, pre-existing Arabic idiom, which is a fixed expression. The difference in diacritics (يُطبّقون vs. يطبقون) is minor and does not change the meaning.
NMT Performance	Modern NMT systems would handle this very well. “Law of the jungle” is a common idiom with a direct equivalent in its training data. It would almost certainly output قانون الغاب correctly. The fluency and accuracy of both human translations are what a good NMT engine aims to produce.

Table 26. Example 17: “Ladies and gentlemen, lend me your ears”

Theory	Analysis
Nida's Dynamic Equivalence	Both translations are dynamically equivalent. They correctly translate the address سيداتي وسادتي and the core meaning استمعوا (listen).
Vinay & Darbelnet's Strategies	The address "Ladies and gentlemen" is a candidate for a Literal Translation, but the standardized equivalent سيداتي وسادتي is so entrenched it acts as an Equivalence. The verb phrase "lend me your ears" cannot be translated literally. It requires Equivalence. اصغوا اليّ is a stronger, more established equivalence for this Shakespearean plea, while استمعوا اليّ is a more general, direct modulation.
NMT Performance	NMT would correctly handle the address. For the verb, it might have a higher probability of choosing the more common استمعوا, but with sufficient context (e.g., the rest of the speech), it could also generate the more formal اصغوا. The choice would depend on the formality of the source text's register in its training data.

Table 27. Example 18: "He serves two masters"

Theory	Analysis
Nida's Dynamic Equivalence	Both translations are fully equivalent. The idiom "serve two masters" (meaning to be loyal to two conflicting parties) is perfectly captured by the Arabic يخدم سيدين. The clarifying phrase في آن واحد (at the same time) is necessary in Arabic to convey the inherent conflict of the idiom, making the meaning immediately clear to the target audience.
Vinay & Darbelnet's Strategies	This is primarily Equivalence for the idiom. The addition of في آن واحد is an example of Explicitation (making implicit information explicit), which is one of Vinay & Darbelnet's procedures. The inclusion of the pronoun هو (he) in Translation 2 is a stylistic choice (Amplification) common in Arabic for emphasis or flow, but it doesn't change the core meaning.
NMT Performance	This is a good test for NMT. A basic word-for-word translation would fail ("هو يخدم سيدين" misses the conflict). However, a sophisticated NMT model, trained on vast datasets, would likely recognize this as a common idiom and generate the full, correct phrase يخدم سيدين في آن واحد, understanding that the explicitation is required for accurate meaning in Arabic.

Table 28. Example 19: "Let us turn a new page"

Theory	Analysis
Nida's Dynamic Equivalence	Both achieve dynamic equivalence perfectly. They use the standard Arabic metaphor for starting anew. The choice of verb (افتح - to open vs. اقلب - to turn) is a minor stylistic variation within the same cultural metaphor. Both sound natural and evoke the same sense of a fresh start.
Vinay & Darbelnet	This is another clear case of Equivalence. The metaphor of a "new page" or "new leaf" is common in both cultures. The translators are not translating the words "turn" or "open" literally; they are selecting the most natural verb that collocates with اصفحة in Arabic to express the same idea.
NMT Performance	NMT would excel here. The metaphor is universal and has direct equivalents. It would confidently output either افتح صفحة جديدة or اقلب صفحة جديدة (or a close variant), as both are highly fluent and correct. The choice might be based on the most frequent pairing in its parallel corpora.

Table 29. Example 20: "His name was on the tip of my tongue"

Theory	Analysis
Nida's Dynamic Equivalence	This is a perfect example of a successful Calque (a literal translation that becomes an accepted idiom). The English idiom maps exactly onto the Arabic word-for-word: on the tip of my tongue = على طرف لساني. The effect on the target reader is identical: immediate recognition of the feeling of almost remembering something. Ghazala (2008) translates the idiom as "كان اسمه على راس لساني" (p. 129)

Theory	Analysis
Vinay & Darbelnet's Strategies	This is a Literal Translation that functions as a perfect Equivalence. It is one of the rare cases where a word-for-word translation does not distort meaning and is the officially recognized idiom in the target language. The strategy is literal, but the result is equivalency.
NMT Performance	This is the simplest case for an NMT system. Since the translation is literal and the structures align perfectly, any competent NMT engine would produce this translation with 100% accuracy and fluency. It would be a direct, unambiguous mapping.