A Comparative Linguistic Analysis Of Human And AI Medical Term Translations Into Arabic

Received 2025-04-08
Accepted 2025-10-11
Published 2025-12-27

Ahmad Ali Assiri
Department of General Courses, Faculty of Arts and Humanities, King Abdulaziz University, Saudi Arabia
aassiri1@kau.edu.sa

To cite this article: Assiri, Ahmad Ali Assiri. (2026). A Comparative Linguistic Analysis Of Human And AI Medical Term Translations Into Arabic. Ijaz Arabi: Journal of Arabic Learning, 9 (1), 11-28, DOI: https://doi.org/10.18860 /ijazarabi.V9i1.32531

Abstract

In recent years, artificial intelligence has become a dominant force in language translation, even in high-stakes fields such as medicine. Nevertheless, questions persist on the accuracy and context-specific validity of AI-provided translations of medical terms. This study aims to evaluate the quality of AI translations in comparison to those of human professionals, focusing on linguistic accuracy, clinical appropriateness, and adherence to medical discourse norms. It specifically analyzes the extent to which AI tools, such as ChatGPT-4, successfully transcribe English medical terminologies into Arabic, identifies recurrent linguistic difficulties faced in AI translation, and discuss possible avenues to enhance the quality of translation through AI. The study draws its conceptual foundation from Halliday's Ideational Functional Linguistics (SFL) approach and focuses on the ideational, interpersonal, and textual functions of medical language. The comparative qualitative analysis was undertaken using purposive sample selection of ten English medical terms from various clinical subfields. Each of these terms was translated by both AI and human translators and analyzed using a descriptive-analytical approach, taking into consideration term accuracy, syntagmatic structure, grammatical accuracy, and register appropriateness. The findings reveal that AI translations are structurally fluent but often lack semantic accuracy, subject-specific terminological usage, and register appropriateness. The comparison shows that human translations are invariably superior in terms of appropriateness to Arabic clinical conventions, particularly in procedural-to-diagnostic contexts and descriptions of pathophysiology. The current study concludes that AI translation tools demand substantial enhancement in accordance with exposure to specialized Arabic medical corpora, enhanced genre sensibility, and post-editing measures. These findings have significant implications for the integration of AI in healthcare communication and support the strategic objectives of Saudi Arabia's Vision 2030 in advancing AI applications in medicine. The study contributes to the broader discourse on the responsible use of AI in sensitive domains, advocating for hybrid translation models that combine machine efficiency with human linguistic expertise.

Keywords: AI; Medical; Translation; Linguistics; Saudi Arabia's Vision 2030

INTRODUCTION

The application of artificial intelligence (AI) in the world of language translation has presented positive prospects to counter the challenges of translating complex and highly technical texts with particular reference to the medical world. The concomitant development of the type of neural machine translation (NMT) models with extensive

language models such as ChatGPT-4 has evoked additional research into the role of AI systems in mediating communication in healthcare environments (Stahlberg, 2020; Ponnusamy, et al., 2022; Sun & Zhou, 2023; Ye, 2024). While the systems prove to be competent in performing generic language translations with enhanced fluency, their effectiveness in technical knowledge fields with particular reference to medical terminologies in linguistically diverse and morphologically complex language varieties such as Arabic necessitategreater exploration and discussion (Roba, 2023; Elhadary, 2023).

The accuracy of medical translations is extremely important since any misunderstanding or imprecision might result in inappropriate treatments and compromised patient safety (Karliner et al., 2007; Banner et al., 2019; Leonardi, 2022; Xiaoqing Mao & Thakkar, 2023; Nazi & Peng, 2024). Traditionally, this role has fallen to skilled human translators with medical as well as linguistics skills (Azizova, 2023). However, with the advancement in the capabilities of artificial intelligence (AI), there is an imperative to ascertain the accuracy to which machine translation can match the high standards of accuracy, sensitivity, and contextual appropriateness required in clinical environments. Dahal and Aoun (2023) recognize the potential of AI to be utilized to improve the availability of health information to multilingual populations but also point to its shortcoming in dealing with specialist language and cultural nuances. Similarly, Nazi and Peng (2024) highlight the efficiency of artificial intelligence for handling clinical terminology but acknowledge notable limitations in real-world healthcare environments stemming from accuracy and reliability issues.

The challenges are especially marked with regard to the Arabic language, which has a diglossic nature, inflectional morphology, and high syntactic variability. The complexities involved in translating medical literature into Arabic have been widely documented in recent literature (Yaseen, 2013; Al-Jarf, 2018; Olimat, 2019; Almahasees & Husienat, 2024), highlighting cases where a single English word has multiple equivalents in Arabic, creating inconsistencies and potential miscommunications in medical contexts. Moreover, Roba (2023), Elhadary (2023), and Nasution and Onan (2024) argue that the highly complex grammatical structure and cultural sensitivities intrinsic to the Arabic language are a strong basis for challenging the adequacy of AI models trained mostly on English or Romance languages.

Systemic Functional Linguistics offers a comprehensive model of language as a social semiotic system (Halliday, 1985). It is particularly relevant to the consideration of language in naturalistic contexts in such areas as the translation of medical vocabulary, in which the accuracy and appropriateness of information to context are paramount. The model explains language in terms of three key functions: ideational, interpersonal, and textual, thus facilitating an increased understanding of meaning-making in language as well as language interaction with context (Halliday, 1985; Halliday & Matthiessen, 2013).

1. Ideational function; in its study of the ideational function, this work evaluates the way different processes, agents, and conditions are represented in human and artificial intelligence translations. A focus of this analysis is representations of actions such as treatments and diagnoses along with key agents such as medical professionals and patients. Knowledge of these representations is significant in its usefulness in judging the effectiveness with which medical actions and roles are transmitted by different agents. The work also covers the use of technical vocabulary as well as the accuracy of term choices. The purpose of this inquiry is to examine the extent to which human

and AI translators approach medical discourse with fidelity in regard to specialist terms.

2. Interpersonal function; the analysis employs modality and attitude in its investigation of modal verbs with additional implication-carrier markers of likeliness, obligation, and suggestion in medical directives. This role shows the way in which translations engage the communicative properties of medical texts to guide patient action and knowledge. To examine the engagement and stance in relation to the audience and ascertain the extent to which translations engage the reader while preserving the instructive and directive purpose of medical texts, the interaction and stance with respect to the audience are also considered.
3. Textual function; cohesion and coherence are essential areas of examination in the present study that measures the way in which translations relate concepts and maintain a logical flow. The study systematically reviews the effectiveness of such translations to describe complex medical processes in a transparent and coherent way. Theme and informational structures are subjected to critical analysis to identify the way in which information is prioritized and presented in the translations. This aspect is important in understanding the extent to which the text supports the reader in digesting complex medical information. Detailed qualitative analysis is carried out on selected texts with a careful comparison of the strategies used by different translations in coping with complex medical information, with a particular focus on sensitivity to meaning, tone, and register. This in-depth analysis supports the determination of minute differences and flaws in AI-based translations.

The significance of studying the effectiveness of artificial intelligence in medical translation is highlighted by national plans such as Saudi Arabia's Vision 2030, which promotes the implementation of advanced technology in strategic sectors, including healthcare. Pursuing this strategic transformation, the quality of medical care while maximizing efficiency and availability through AI is seen as an essential priority. For this reason, an understanding of the limits and capabilities of AI in translating medical content into Arabic closely aligns with Saudi Arabia's developmental strategies and the wider vision of elevating healthcare provision through innovative strategies.

While existing research provides valuable insights into the general effectiveness of AI translation and the dynamics between human and AI collaborators in the context of healthcare communication, an identifiable gap remains in comparative studies that solely evaluate both AI and human translations of Arabic medical terms. Most previous research, including that of Azizova (2023), Al-Jarf (2018), and Dahal and Aoun (2023), has a tendency to either address translation problems in a range of language pairs or focus on the educational needs and complexities of working with AI systems. Nevertheless, a significant lack of in-depth linguistic assessments systematically and extensively comparing human and AI translations of Arabic medical terms can be observed. To bridge this gap, the current study offers a comparative linguistic analysis of human and AI translation of English medical terms into Arabic. This research draws on Halliday’s (1985) Systemic Functional Linguistics (SFL) as its foundation, evaluating translations based on their ideational, interpersonal, and textual functions. In so doing, it identifies particular areas where AI systems can be limited or show proficiency, thereby contributing to the development of more accurate, register-sensitive, and contextually relevant tools for AI translation.

In essence, this research project aims to test the reliability of artificial intelligence in an area in which accuracy in translation is critical and small linguistic details can have substantive effects on patient care. Conducting this study increases theoretical knowledge and practical improvements and offers worthwhile viewpoints to health practitioners, translators, policymakers, and creators of AI language software. The implications go beyond simple linguistic analysis because it supports strategic initiatives regarding the digital transformation of healthcare with the added moral obligation to promote safe and effective communication in multilingual settings of healthcare. This study aims to answer how accurate are AI-generated translations of medical terminology compared to those produced by human experts in Arabic? What are the main linguistic challenges AI translation tools encounter when translating medical terms from English into Arabic? How can AI translation tools be improved to better handle the translation of medical terminology into Arabic?

METHOD

The study utilizes comparative cross-sectional analysis to evaluate and compare the quality of medical term translations by human experts with those delivered by machine-based technology. This approach allows an in-depth analysis of translation quality that also identifies the nuances and complexities involved in medical term translations. The corpus of this study consisted of medical statements and segments of diagnoses derived from authentic medical reports representing different specialties, such as cardiology, pulmonology, hematology, pathology, surgery, and emergency medicine. The latter were intentionally included to ensure that the study encompassed a wide range of medical terminology and contextual uses. Overall, ten medical terms were chosen according to three basic criteria: (1) these terms are commonly used as diagnostic and procedural terms in medical literature; (2) these terms are well-documented to pose difficulties when translated into Arabic due to morphological, syntagmatic, or semantic complexities; and (3) their meanings depend on context, requiring careful interpretation to be precisely conveyed with meaning and proper register in Arabic.

The selected terms were subjected to analysis in the context of their respective sentences in the original literature in medicine with a focus on evaluating not only the translation of each term individually but also the syntagmatic, semantic, and discursive choices apparent in their embodied usage. All ten terms were translated into Arabic through applying two approaches: (1) AI-produced translations were accessed through the subscription plan of ChatGPT-4 to ensure exposure to the full capacity of the model during data gathering; and (2) human translations were carried out by an experienced translator who possesses a PhD in Translation Studies and has had in excess of ten years of experience translating in the medical context. This dual-method strategy ensured a comprehensive comparative linguistic analysis across the three systemic functional variables: ideational, interpersonal, and textual.

The study undertook a comparative qualitative analysis of ten carefully selected English medical terms in authentic medical report contexts. Each term was analyzed in translations carried out by artificial intelligence as well as humans. The focus of comparison was on contextual translations of the terms as opposed to their isolated forms in order to examine the construction and retention of meaning. This research was based in the theoretical context of Systemic Functional Linguistics (SFL) and outlined three areas in which discussion should be explored in greater depth. Each set of translations

DOI: 10.18860 /ijazarabi.V9i1.32531
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

was subjected to descriptive-analytical examination with a focus on factors such as terminological accuracy, syntactic structure, grammaticality, and register appropriateness with all areas of investigation explored in some depth. To avoid using coding or frequency analysis, the study was based on extensive qualitative discussion centered on individual items with supplementary context-based rationale. The outcome of this approach was a sequence of detailed comparative observations on the linguistic, terminological, and discursive strengths and weaknesses of human and AI translations with specific focus on the way in which AI systems cope with the vagaries of Arabic medical terminology.

RESULTS AND DISCUSSION

Table 1. Translation of Cardiomegaly

Medical Term	Example	AI Translation	Human Translation
Cardiomegaly	Cardiomegaly (540 g) with mild biventricular dilatation	تضخم القلب (٥٤٠ غرامًا) مع توسع بسيط في البطينين	تضخم في عضلة القلب (٥٤٠ غرامًا) مع توسع طفيف في البطينين

Table (1) shows the clinical term, cardiomegaly, used in a diagnostic context, along with its degree and anatomical connotation. The differences in linguistic variations between the human and AI translations are considerable and in line with the aims of this study. In the AI translation, تَضخم القلب (٥٤٠ غرامًا) مع توسع بسيط في البطينين, the term "cardiomegaly" is translated as تَضخُّم القلب, a term often used in colloquial Arabic contexts but one lacking the required accuracy in medical reports. By contrast, the human translation, تضخم في عضلة القلب., better conveys the intended term by adding عضلة القلب (cardiac muscle), thus ensuring fidelity to the anatomical locator denoted by the term "cardiomegaly," instead of implying an overall heart enlargement. This reveals the superior semantic accuracy of the human translation, with particular attention to its ability to express the clinical accuracy required in medical language.

The AI translation shows a tendency toward lexical generalization, with a preference for the word القلب (heart) instead of the more technical عضلة القلب (myocardium). This reflects a difficulty in properly distinguishing between general and technical registers, a major challenge in the field of Arabic medical translation where the choice of terminology can have significant diagnostic implications. This finding aligns with previous research (e.g., Al-Jarf, 2018; Roba, 2023), which stresses that Arabic medical terms often require disambiguation that AI systems struggle to provide. This example shows the need for AI models to be trained on specialized corpora that emphasize correct terminological norms in Arabic. Enhancing AI's capacity to differentiate between lay and clinical terms is crucial for advancing translation accuracy in high-stakes fields like medicine.

Ideationally speaking, the human translation better represents the underlying processes (تضخم) and actors (عضلة القلب) involved. While the AI translation conforms to grammatical conventions, it does not capture the anatomically specific character of the initial terms and thus might compromise on the identification of the heart muscle as an individual clinical entity. This deviation affects the degree to which experiential meaning in this case, the medical ailment can be transferred. Both translations take a modality-neutral approach since modality has successfully been neutralized in order to conform to report standards. However, the human translation has a slightly more formal register

DOI: 10.18860 /ijazarabi.V9i1.32531
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

(طفيف vs. بسيط) in congruence with standard conventions in Arab medicine, thus improving its professional utility and clinical usefulness. Cohesive ties are maintained in both renditions; however, the human rendition shows a higher level of consistency in its lexis with standard Arab medical vocabulary. The systematic ordering of information, from the illness to the extent and then to the involved anatomy, is organized in a way that better conforms to Arab medical conventions.

Table 2. Translation Of Arteriosclerotic Heart Disease

Medical Term	Example	AI Translation	Human Translation
Arteriosclerotic heart disease	Arteriosclerotic heart disease, multifocal, severe	مرض القلب التصلبي الشرياني، متعدد البؤروشديد	مرض القلب الناتج عن التصلب الشرياني، شديد ومتعدد البؤر

The term "arteriosclerotic heart disease" involves a multifaceted designation with modifiers to indicate distribution (multifocal) and severity (severe). Each translation attempt aims to preserve these components in the Arabic language; however, significant differences emerge regarding their effectiveness in conveying the technical and functional connotations of the original term. The AI-provided translation مرض القلب التصلبي الشرياني، متعدد البؤروشديد shows a simple lexical rendering; however, it shows formalistic word order and clumsy syntax. For example, the placement of "وشديد" at the end distorts the grammatical flow inherent in standard Arabic sentence syntax. More importantly التصلبي الشرياني although seeming technical in its individual parts, suffers from syntactical imprecision since it constitutes a compressed noun sequence lacking the connecting elaboration typical of medical Arabic. By contrast, the human translation مرض القلب الناتج عن التصلب الشرياني، شديد ومتعدد البؤرenjoys semantic clarity and coherence with greater effectiveness. The phrase الناتج عن التصلب الشرياني conveys causality and etiology effectively and thus improves comprehensibility while conforming to conventions in Arabic medical language. Moreover, the ordering of شديد ومتعدد البؤر conforms to Arabic conventions in enumeration of the adjective and thus maximizes readability while maintaining correct diagnosis accuracy.

The AI poses a particular challenge because it analyzes "arteriosclerotic heart disease" as a string of compounds and doesn't explain the correlation between the disease and its causes. This is because the AI has problems in dealing with noun-noun collocations effectively and converting them well into natural clause-based expressions in Arabic; this is a typical challenge one would encounter in translating specialist literature into morphologically complex languages like Arabic (cf. Roba, 2023; Al-Jarf, 2018). Second, the adjective ordering in متعدد البؤروشديد violates the expected collocation conventions in Arabic, in which distribution descriptors are generally placed in front of descriptors of quality relating to severity. This example shows that it requires the AI translations to include the capacity to perform syntactic re-arrangement in accordance with the Arabic's subject-object-verb pattern and preferred noun-adjective structures. Second, the AI would be helped if it received wider training on Arab medical corpora to better grasp collocation patterns in addition to clinical terminology and in etiological descriptions in medical contexts.

The human translation better realizes experiential meaning by shifting from the noun phrase to the relational clause الناتج عن التصلب الشرياني thus successfully capturing the

etiology of cardiovascular disease. By contrast with the AI translation, which reads as correct but uses a short structure tending to subvert ideational purpose on grounds of insufficive clarity in representing relationships between concepts (disease, cause, extent, and severity), the human translation employs a formal and explanatory style in line with medical discourse conventions. The use of the phrase "**شديد ومتعدد البؤر**" is also in line with clinical vocabulary widespread in the Arabic language, avoiding the prospective semantic imprecision or stylistic contortion involved. Besides, the human translation exhibits better text cohesion and logical flow of information. The disease > cause > extent/distribution sequencing conforms to a traditional planning schema in medical reports found in Arabic. In contrast with this natural sequencing in a clinical context, the word ordering in the AI renders thematic coherence impossible to achieve, with deleterious consequences to reader understanding.

Table 3. Translation of Pulmonary Edema

Medical Term	Example	AI Translation	Human Translation
Pulmonary edema	"The pulmonary parenchyma is diffusely congested and edematous"	نسيج الرئة مُحتقن ومنتفخ بشكل منتشر	نسيج الرئة البرنشيمي مُحتقن ومُتوذم بشكل معمّم

The diagnostic statement encapsulates a description of a pathological state of the lungs with anatomical specificity (pulmonary parenchyma), and extent and character of the pathology (congested and edematous), to both of which it gives prominence. Both Arabic translations show considerable differences in terminological precision, linguistic register, and contextual appropriateness. The AI translation **نسيج الرئة مُحقتن ومُنتفخ بشكل منتشر** uses the generic term **منتفخ** (swollen), which falls short of clinical precision required by the term edematous. Further, **نسيج الرئة** (lung tissue) is correct but does not convey the anatomical specificity entailed in the original phrase pulmonary parenchyma. While the AI translation increases the language accessibility to some extent, it does so to the expense of clinical precision required in medical narration. However, the human translation **نسيج الرئة البرنشيمي مُحقتن ومُتوذم بشكل معمّم** shows an upgraded level of terminological accuracy. The term **البرنشيمي** (parenchymal) is a direct equivalent to the original technical term and hence preserves consistency with standardized Arabic medical vocabulary. Besides, **متوذم** is an exact equivalent of edematous and thus retains the pathology context. The phrase **بشكل معمّم** is a more natural and medically fitting rendition of diffusely than the AI version **بشكل نشر**. This example demonstrates key challenges to AI technology like ChatGPT-4 in translating technical medical content. A key issue has to do with the choice of technical terminology, in which the AI has a preference for colloquial vocabulary like **منتفخ** over more clinically accurate options like **متوذم** that can compromise clarity in a professional context.

Second, the translations themselves lack anatomical accuracy based on the exclusion of complex terminology like **البرنشيمي** due to their low frequency in generalized training corpuses. A final issue is one of register appropriateness: although the translated text is adequately clear, it is inappropriate in a formal register generally expected in medical writing. Overall, these issues serve to emphasize weaknesses in AI to follow domain-specific criteria and sensitivity to registers in translating medical descriptions into Arabic. To correct these shortcomings, it is essential that AI tools are trained on corpora

of Arabic medical content that expose them to standard professional vocabulary. Additionally, AI tools should be set to detect conventions specific to registers, including the recognition of situations in which formal clinical vocabulary is required versus straightforward vocabulary.

The experiential connotation related to the term pulmonary parenchyma, congested, and edematous is better maintained in the human translation. The phrase **البرنشيمي** and **متوذم** precisely capture the distinct referents and criteria established in the original document. In contrast, the AI version settles into a generic interpretation that threatens compromise of the semantic accuracy of the medical diagnosis. The human translation conforms to the expected tone used in medical literature and thus ensures objectivity as well as technical accuracy. The use of **منتفخ** by the AI degrades the formal character and inadvertently makes the diagnosis less credible. This shift in register can undermine the authoritativeness and gravitas required of medical reports. The human translation enjoys superior cohesion of text as well as structural clarity. The structural pattern (anatomical term → condition 1 → condition 2) mirrors the structural pattern found in the original document. TheBushraing **بشكل معمّم** successfully achieves a natural flow and stress pattern while the use of**ιvε شدن بشكل منتشر** in the AI version reads linguistically flat and inappropriate to technical writing.

Table 4. **Translation of Contusion**

Medical Term	Example	AI Translation	Human Translation
Contusion	Multiple mentions such as “purple contusion,” “faint pink contusion,” and “abraded red-black-purple contusion”	عدة إشارات مثل: "رضة أرجوانية"، و"رضة وردية باهتة"، و"رضة مُجروحة حمراء-سوداء-أرجوانية	عدة إشارات مثل: كدمة أرجوانية، كدمة وردية باهتة، وكدمة مُجروحة بألوان حمراء وسوداء وأرجوانية

The sentence in Table (4) above comprises an elaborate list of the different types of injuries noted in clinical and forensic situations. The word "contusion" is used with varying qualifiers that denote surface and colour. The effectiveness of the translation depends on the accurate use of medical vocabulary, clear identification of colours, and a syntactically coherent final ordering of the listed items. The AI-produced translation employs the word **رضة** to refer to the word "contusion," which is although common in everyday situations is not commonly found in the formal medical register. In contrast, the human-rendered translation prefers using **كدمة**, which is the conventional and formal term used in medical Arabic environments. In addition, the AI's use of colours in hyphenated syntax (**حمراء-سوداء-أرجوانية**) imitates English punctuation to create a construction that feels unidiomatic in Arabic syntactic rules. The human translation corrects this syntactically jarring construction and increases clarity and linguistic consistency by inserting **بألوان** before the list of colours.

This example explains a number of serious issues related to AI translations of medical terms in Arabic. First, the preference for **رضة** over **كدمة** by the AI demonstrates terminological incompatibility with clinical practice. In addition, the AI struggles to produce coherent lists and use proper punctuation in Arabic, leading to structural issues within the translated documents. Finally, while the intended meaning is preserved, the stylistic register defaults towards colloquial rather than formal professional discourse that is crucial in medical or forensic reporting. Together, these shortcomings demonstrate a

DOI: 10.18860 /ijazarabi.V9i1.32531
ISSN(print): 2620-5912 | ISSN(online):2620-5947
ejournal.uin-malang.ac.id/index.php/ijazarabi/index

broader limitation in AI systems for their generic lexical and syntactic choices, even in specialized areas like injury documentation and morbid description. The noted limitations show that there is a need for improvement in translation software to ensure greater accommodation of preferred terminologies within the discipline, particularly in forensic or medical reporting. The inclusion of Arabic corpus data from medical and judicial sources would support AI's ability to conform to appropriate registers and terminologies. Finally, polishing punctuation conventions and phrase structure restrictions in Arabic would improve both syntactic fluency and general readability.

The human translation better communicates the experiential meaning of the original. By using the term كدمة, it is adjusted to fit the standards of medical language more closely. The phrase بألوان حمراء وسوداء وأرجوانية better conveys the complexity of the described injuries than the hyphenated listing done by the AI. Moreover, the human translation conveys the shades in appearance (e.g., مُجروحة) in a smoother way. Despite an attempt to be objective in both translations, the human one has a sense of greater gravity and precision, traits expected in medical or post-mortem reports. The almost-informality and less standard words used in the AI one somewhat take away from its professional tone. The human translation shows better cohesion and coherence in addition. The repetitive use of كدمة ensures coherence and preserves the structure of the original piece. In contrast to this, the sporadic punctuation and varied structural components in the AI one undermine the desired flow and continuity found in the work of a professionally written report.

Table 5. Translation of Petechiae

Medical Term	Example	AI Translation	Human Translation
Petechiae	No facial, oral mucosal, or conjunctival petechiae	لا توجد نزيفات نقطية في الوجه أو الغشاء المخاطي للفم أو الملتحمة	لا توجد نمشات نزفية في الوجه أو الغشاء المخاطي الفموي أو الملتحمة

The statement in Table (5) above illustrates an idiomatic medical formula commonly encountered in clinical or forensic reports. It identifies the absence of a certain pathological sign (petechiae) on three anatomical locations. Accuracy in vocabulary and use of contextually specific lexis are crucial in communicating both the pathologically specific sense and formal register of the original text. The AI translation "لا توجد نزيفات نقطية في الوجه أو الغشاء المخاطي للفم أو الملتحمة" is semantically accurate; however, it is lacking in terminologically correct vocabulary. نزيفات نقطية is a descriptive translation of petechiae (literally: pinpoint hemorrhages), which might be suitable to use with a layman but less used in clinical Arabic medical discourse. The human translation لا توجد نمشات نزفية في الوجه أو الغشاء المخاطي الفموي أو الملتحمة correctly employs نمشات نزفية as an accepted term in Arabic clinical writing to describe petechiae, thus being more professional in a diagnostic context. Again, the phrase الغشاء المخاطي الفموي shows a more clinically reasonable construction than the phrase الغشاء المخاطي للفم in accordance with the collocation conventions of medical writing in Arabic. The translation shows an exemplar of AI's inability to prioritize domain-specific terminology. Rather than resorting to the standard medical term نمشات نزفية, the AI opts for a descriptive take that, although semantically correct, lacks professional depth. The addition of لفم in the AI's phrase (للفم) also indicates dependency on syntactically generic templates rather than on conforming to conventions

of medical phrase construction in vocabulary to include the small but significant anatomically specific determinatives in medical vocab. This indicates a greater issue in AI's abilities to negotiate clinically relevant vocabulary and Arabic compound noun construction in clinical texts.

This example supports the importance of developing artificial intelligence systems by adding specialized medical Arabic corpora that prioritize formal, internationally accepted terminology. AI models should be equipped with the capacity to go beyond simple understanding of the literal sense; they should be able to recognize contextually appropriate synonyms with which clinical practitioners would be familiar. The effectiveness of AI in effectively translating such highly specialized words as petechiae depends on its ability to distinguish between usage in everyday language and clinical use and to express this in real-time formulations appropriately.

The experiential explanation of the sentence shows the lack of a specific clinical sign in different areas. The use of نمشات نزفية in the human translation is better suited to conveying this implication since it has an immediate reference to the established medical term. By contrast, the use of نزيفات نقطية by the AI introduces an interpretive element liable to cause clinical or forensic confusion. Both translations are neutral in tone and factual in their character in accordance with clinical reporting. However, the vocabulary in the human translation is preferable to the professional tone expected in medical discourse. While grammatically correct, the AI's usage is less assertive in their clinical context and is liable to affect their perceived credibility. The human translation also has greater textual coherence through grammatical pattern standardization in all areas of anatomy. The AI text fails to accomplish this. In addition, the addition of the adjective الفموي in the human translation allows for a more coherent syntactic flow and is in greater accord with conventional Arabic medical discourse conventions.

Table 6. Translation of Sickle Cell Trait

Medical Term	Example	AI Translation	Human Translation
Sickle cell trait	"This quantitative result is indicative of sickle cell trait" – under Comments in the autopsy findings	تشير هذه النتيجة الكمية إلى وجود سمة الخلية المنجلية	تشير هذه النتيجة الكمية إلى وجود سمة الخلية المنجلية (حامل لمرض الخلية المنجلية)

The statement in Table (6) contemplates a formal data-driven role in the fields of pathology and forensic reporting, specifically in the context of laboratory findings explanation. It is essential that the Arabic term explains the genetic status clearly in order to avoid confusion with its severe counterpart, sickle cell anemia, with consideration to both clinical accuracy and term usage in historical context. The AI-proposed translation to تشيرهذه النتيجة الكمية إلى وجود سمة الخيلة المنجلية is linguistically accurate and clinically relevant. This translation uses the term سمة الخيلة المنجلية as known within the Arab vocabulary equivalent to sickle cell trait. However, this translation would incidentally prove to be confusing to lay readers who are not equipped with the knowledge to distinguish between the trait (carrier state) and related clinical illness (anemia). By contrast, the human translation "تشير هذه النتيجة الكمية إلى وجود سمة الخلية المنجلية (حامل لمرض الخلية المنجلية)." adequately solves this issue by adding the explanatory parenthesis "(حامل لمرض الخلية المنجلية)." This extra information not only explains the initial sentence but also sustains the meaning of the message with added interpretive assistance, which is particularly needed in forensic

writing such as autopsies in which accuracy in genetic and diagnotic vocabulary is crucial. This illustration shows that while artificial intelligence might be technically correct in its translation, it can be insensitive to context. The word سمة الخلية المنجلية is correct; however, it assumes that the reader will interpret it correctly, which is a significant hazard in cross-disciplinary writing such as autopsy reports. AI lacks an element of discursive flexibility to anticipate whether an explanation or parenthetical clarification would be needed, particularly with regard to genetic vocabulary, which can be lacking in identical clarity in Arabic.

This indicates a broader problem: AI often falls into a simplistic one-to-one correspondence model with an insufficient consideration of the reader's perspective or the accuracy of interpretive conclusions, both of which are crucial in medical and forensic reports. To compensate for this weakness, the AI systems would be required to include interpretive pointers and qualified language in cases in which medical terminology might have multiple connotations (e.g., trait versus disease). Having domain-specific post-editing procedures or teaching AI models to recognize cases in which clarification would be clinically or legally relevant would significantly improve the quality of the output. Both translations attempt to present an analysis of quantitative data in relation to a medical-genetic deficiency. The AI translation focuses on the experiential role of the content, while the human translation focuses on the ideational role by explaining the relevance of the quality in the context of a carrier—a stylistic approach better suited to the communication demands of an autopsy report. The human translation increases reader engagement and accuracy of interpretation through the proactive clarification of areas of confusion and bringing them into focus. The AI translation makes use of formal and abstract phraseology and fails to capture the attitudinal nuances that would avoid confusion with respect to the assertion. The human translation aids in coherence and readability through parenthetical explanation. This process better prepares the reader to accept the intended meaning of the term in accord with the standards of forensic linguistic accuracy, where the absence of ambiguity is of paramount consideration.

Table 7. Translation of Thoracostomy

Medical Term	Example	AI Translation	Human Translation
Thoracostomy	“Thoracostomy incision (3.6 cm...)” and again on the left chest	شق جراحي لتصريف الصدر (٣,٦ سم...)" ومرة أخرى في الجانب الأيسر من الصدر	شق صدري لتصريف الهواء أو السوائل (٣,٦ سم...)، وآخر مماثل في الجانب الأيسر من الصدر

The term "thoracostomy" connotes both operative and postmortem documentation procedures. Thoracostomy in this context is an incision in the chest wall to drain pleural contents, which are possibly air, fluids, or blood. The correct Arabic translation should maintain the technical accuracy and procedural meaning inherent in the original term. The AI translation of "شق جراحي لتصريف الصدر (٣,٦ سم.) ومرة أخرى في الجانب الأيسر من الصدر" is grammatically correct; however, it uses the imprecise and colloquial term "شق جراحي لتصريف الصدر" which might imply evacuating the whole chest instead of precisely dealing with the pleural contents (blood or air). For this reason, the clinical purpose of thoracostomy in this translation is lost because it is based on terminologically vague language. The human translation "شق صدري لتصريف الهواء أو السوائل (٣,٦ سم.)، وآخر مماثل في الجانب الأيسر من الصدر" clearly indicates both procedural purpose and medical accuracy required to clearly translate. A better fit into the term "thoracostomy" is the term "شق صدري" and adding the phrase "لتصريف

"الهواء أو السوائل" offers essential functional explanation to secure conformance with clinically accepted standards of documentation. The human translation accommodates better semantic accuracy and procedural appropriateness.

This instance evinces the challenges faced by artificial intelligence in effectively translating specialized medical procedure terminology. The AI system recognizes "thoracostomy" as a generic surgical cut (شق جراحي) and tries to express it through an inexact phrase (لتصريف الصدر), with no anatomical accuracy and clinical value. In addition to this, the AI fails to express the implied repetition of the process (again on the left chest) with the same recognition of clinical formality and accuracy that is present in human language (وآخرممائل في الجانب الأيسرمن الصدر) and instead resorts to the colloquial and somewhat indefinite word مرة أخرى.

This example demonstrates the need for artificial intelligence models to develop a more sophisticated medical procedural vocabulary, specifically in the Arabic language, where compound phrases and contextually dependent expressions are common. AI models need to be trained to precisely map technical terms to their functionally and terminologically correct equivalents, instead of approximate translations. The addition of structured medical glossaries in combination with procedural corpora would greatly improve AI performance, particularly in the fields of surgical and autopsy reporting.
The human translation reflects a better realization of the experiential elements by specifying the kind of process (شق صدري) relevant to the intended medical purpose (تصريف الهواء أو السوائل). In this respect, the AI translation has a lower level of accuracy; therefore, the ideational function of explaining the actions taken and the reasons behind them is only fulfilled to some extent. The human translation retains professional and clinical discourse quality conforming to medical report standards. In contrast, the AI version, by introducing the phrase مرّة أخرى, has a relatively non-formal style that is inappropriate in the context of surgery reports. Such a fine-grained difference can influence the perceived validity of the text. The human translation gains increased effectiveness with regard to textual cohesion and coherence. By offering a clear explanation of the sequence—primary happening with recurrence—it ensures narrative clarity. However, the AI version fails to enjoy this structural coherence and thus reads somewhat disjointed.

Table 8. Translation of Intraosseous Catheter

Medical Term	Example	AI Translation	Human Translation
Intraosseous catheter	Intraosseous catheter with attached tubing, right tibia	قسطرة داخل العظم مع أنبوب موصول، في عظمة الساق اليمنى (الظنبوب)	قسطرة داخل نخاع العظم موصولة بأنبوب، في الظنبوب الأيمن

The term "intraosseous catheter" implies the direct insertion of an emergency intravascular access device into the bone marrow, a technique commonly used in critical care and forensic procedures. A correct translation should convey the anatomical accuracy and clinical relevance of the procedure for better understanding within medical or pathology reports. The AI-based translation "قسطرة داخل العظم مع أنبوب موصول، في عظمة الساق اليمنى (الظنبوب)" correctly illustrates the structural components and their specific placement; however, it does not fully address anatomical accuracy. The phrase داخل العظم (inside the bone) is not precise, as it does not clearly convey the technical correctness of intraosseous

as being within the cavity of the bone marrow. Moreover, although عظمة الساق اليمنى is colloquially acceptable in everyday Arabic, it is considered less desirable in medical terminology, where the correct nomenclature is known to be الظنبوب. In contrast, the human translation "قسطرة داخل نخاع العظم موصولة بأنبوب، في الظنبوب الأيمن" successfully utilizes the anatomical term نخاع العظم (bone marrow) while maintaining the equivalent clinical meaning. Additionally, the utilization of الظنبوب الأيمنdemonstrates professional terminology and adherence to medical precision. This example illustrates better adherence to both medical vocabulary and formal norms typical of Arabic medical texts. This example points to a recurring difficulty in the field of AI-powered medical translation: semantic generalization. The artificial intelligence system avoids using specialized terms like نخاع العظم, presumably due to its lack of adequate exposure to technical medical Arabic in its training data. It instead settles on using داخل العظم, a formulation that might be understandable to non-experts but is clinically imprecise. Furthermore, the AI fails to meet expectations with regard to the use of appropriate register, choosing عظمة الساق اليمنى over the more precise الظنبوب الأيمن. This phenomenon points to a broader issue relating to the choice of terminology and stylistic adaptation—essential considerations in the translation of anatomical and procedural information in formal settings, including forensic reports and emergency medicine.

To enhance the quality of artificial intelligence translation in the medical field, it is recommended to use structured bilingual anatomical glossaries in combination with specialized corpora addressing the precise match of Latin-based medical terms and their equivalent Arabic translations. Specifically, it is crucial for AI tools to have the ability to distinguish between casual language and professional jargon, scaling their outputs as required by the target genre in question (e.g., clinical reports, patient education, or forensic reports). The human translation better communicates the experiential value of a medical device placed in the bone marrow in its therapeutic use. In using the term نخاع العظم, it retains the anatomical and procedural fidelity present in the source text. In contrast, the AI translation's use of داخل العظم leads to a loss of technical meaning, potentially threatening clearness in clinical or legal arenas. Both translations maintain the neutral and objective character required in clinical assessment. But the formal architecture and technical vocabulary of the human translation contribute to its professional credibility, an aspect that in the context of high-stakes medical reporting would count significantly. By contrast, the AI translation, with its use of a less technical and colloquial vocabulary (e.g., عظمة الساق اليمنى with the addition of theالظنبوب in parenthesis), weakens the integrity of the reports. The human translation is more cohesive and coherent in its use of the likes of theالظنبوب الأيمن. The AI translation with its addition of the lay and technical terms (e.g., عظمة الساق اليمنى with the addition of the الظنبوب in parenthesis) truncates textual coherence with an inconsistency in register.

Table 9. Translation Of Shortness Of Breath

Medical Term	Example	AI Translation	Human Translation
Shortness of breath	Occasional shortness of breath during physical activity	ضيق تنفس عرضي أثناء النشاط البدني	نوبات متقطعة من ضيق التنفس أثناء النشاط البدني

The clinical description hence outlines a symptom that has an unpredictable pattern in time in addition to a physical stimulus. The translation should retain the medical term (dyspnea) as well as what and in what context it occurs (intermittent, with physical exertion) in formal and precise Arabic. The machine translation ضيق تنفس عرضي أثناء النشاط البدني is grammatically correct and essentially in line with the intended meaning. However, the term ضيق تنفس عرضي might be thought to be overly simplistic or colloquial in the context of Arabic clinical reporting. Although the term عرضي (incidental or occasional) is not strictly incorrect, it is seldom used in medical reporting because it might be vague in situations that require accuracy. In contrast to this, the human-created translation نوبات متقطعة من ضيق التنفس أثناء النشاط البدني.," shows greater stylistic and terminological quality. The addition of نوبات متقطعة (intermittent episodes) conforms to standard medical nomenclature that better indicates the episodic character of such a symptom. This specific phrase is used in standard Arabic medical reports; thus, the human translation meets the accepted standards better in the use of diagnostic terminology.

This example highlights the ongoing struggle of artificial intelligence to master lexical precision and diagnostic terminology. While the word عرضي captures the notion of "occasional," it does not convey the diagnostic specificity inherent in نوبات متقطعة, which is preferred for describing intermittent clinical manifestations. In addition, whereas ضيق تنفس (without the article ال) might be accepted as sufficient in informal settings, the phrase ضيق التنفس is considered more precise in formal medical Arabic contexts. The AI's choice also demonstrates the process of semantic reduction, where the symptom is described using general terms rather than within a carefully constructed diagnostic paradigm, showing AI's weakness in conforming to genre-specific linguistic norms. To enhance the quality of AI-generated translations of symptom descriptions, it is essential to create models that are capable of identifying clinical collocations and formulaic symptom phrases. Such instruments should prioritize terms like نوبة، متقطعة، مزمنة، حادة, and others as core elements of their symptom vocabulary. Fine-tuning AI models through exposure to electronic health record datasets and Arabic medical case studies would greatly advance their ability to mimic authentic clinical discourse.

The human translation better conveys the experiential significance, frequency, and physical stimuli of the symptom. The phrase نوبات متقطعة provides a stronger conceptual basis for explaining the nature of shortness of breath, making it clear that the symptom occurs in recognizable episodes as opposed to being experienced as an indefinite or random event. Further, the human translation maintains formal and objective language in its tone, which is congruent with the clinical jargon expected in medical reports. In contrast, the AI-generated version uses neutral language but is unprofessional in terms of accuracy due to its relatively less accurate language. The human version is better in coherence and clarity, especially in its construction of the noun phrase نوبات متقطعة من ضيق التنفس. This phrasing is congruent with standard medical Arabic, thus enhancing the fluency and understandability of the sentence in a clinical report.

Table 10. Translation of Appendectomy

Medical Term	Example	AI Translation	Human Translation
Appendectomy	Past Medical History: Appendectomy (2005)	التاريخ الطبي السابق: استئصال الزائدة الدودية (2005)	التاريخ المرضي: استئصال الزائدة الدودية (2005)

The term "appendectomy" is used as an identifier in the record of clinical or surgical histories to mark a previous surgical intervention along with its respective date. The translation must maintain medical accuracy, formality of tone, and correct terminology, especially with regard to the expression past medical history and the surgical intervention appendectomy. The AI-generated translation التاريخ الطبي السابق: استئصال الزائدة الدودية (2005)" is both linguistically correct and semantically appropriate. However, it uses the expression التاريخ الطبي السابق, which, although understandable in discourse, is better suited to general discussions of health issues than the specific description of medical conditions, a conventional norm of Arabic medical record-keeping. In contrast, the human-created translation التاريخ المرضي: استئصال الزائدة الدودية (2005) provides a more accurate and accepted term in the field of clinical historiography regarding past medical history. The term للتاريخ المرضي marks the practice by which medical practitioners describe a patient's history with regard to disease, surgery, and relevant conditions in Arabic-speaking healthcare settings. Note that both translations correctly render appendectomy as استئصال الزائدة الدودية, which is the accepted medical terminology.

The use of التاريخ الطبي السابق in machine translation can stem from a literal interpretation of past medical information, which has no discrimination between "medical" as a generic term and "مرضية" (disease-centric) as context-specific. This reflects an even wider issue faced by AI in medical term translation, in which term selection should be informed by conventional usage as opposed to simple semantic approaches. AI processing often shows insensitivity to domain-specific collocations in headings or labels used in clinical recording. While the التاريخ الطبي is not incorrect per se, it does not express the specialized record style expected in Arabic medical reporting in the context of the Middle East. For medical translation efficiency to be enhanced with AI outputs, it is imperative that it be supported by developers with the addition of electronic health record corpuses to training data, thus exposing AI to optimal expressions such as التاريخ المرضي in place of generic terms. AI should be loaded with models that are genre-aware to recognize standard phraseology in medical charts, discharge summaries, or patient histories, in which exactness such as brevity, clarity, and conventionality are paramount. Both translations capture the experiential meaning of an earlier surgical intervention. However, the human translation better preserves the ideational focus of the phrase with respect to medical history related to illness or disease, with a better fit to the original intent. The human version shows better register accuracy in its use of vocabulary and style to reflect a formal level of language found in clinical documentation. By contrast, the AI version reads somewhat wider in its contribution to prosodic pattern in a way that might detract from professional effect in well-prepared clinical documentation. The human translation shows better cohesion in the context of comprehensive Arabic clinical reports. The term التاريخ المرضي seamlessly fits into the hierarchical setup of headings in use in Arab medical

reports. While the AI version is structurally sound, it might read as incongruous or even off-form in comparison to widespread charting conventions.

CONCLUSION

This study sought to evaluate the relative quality of human translator-produced translations versus those produced by AI in a medical language context, specifically from English to Arabic with an in-depth focus on accuracy, contextual appropriateness, and adherence to clinical language conventions. Embedded in a qualitative paradigm grounded in systemic functional linguistics (SFL), the study found that while AI tools like ChatGPT-4 demonstrate significant linguistic mastery and structural cohesion, they continue to struggle with nuanced medical jargon. Among these are the challenges of semantic oversimplification, register insensitivity, terminological inaccuracy, and syntactic unnaturalness—most visible in procedural discourse and pathophysiological term translation. All these are especially severe in Arabic with its highly complex morphological structures, contextually governed phrase usage, and strict genre conventions in clinical reporting. On the other hand, the human translations consistently demonstrated better levels of fidelity to medical terms, contextual appropriateness, and alignment with prevailing norms of professional Arabic discourse. Human translators had a better capacity to understand the clinical implications of terms, utilize domain-specific terminology, and anticipate the communicative needs of the target readership. These strengths were evident across ideational, interpersonal, and textual language functions as analyzed within the SFL model. The findings support the hypothesis that AI-produced translations improve comprehensive understanding but are short of depth required to be used in critical medical and forensic applications. This study thus recommends the development of AI translation technology through the incorporation of systematic Arabic medical corpora with an appreciation of clinical discourse structures as well as post-editing approaches and conforming to standard linguistic parameters. Furthermore, it is suggested that hybrid models that combine the strengths of AI with human oversight are a more reliable way to address the demands of medical communication in situations in which accuracy, clarity, and cultural sensitivity are essential.

In conclusion, this study makes significant contributions to the discussion on the place of artificial intelligence in healthcare communication. The findings have important theoretical as well as practical implications regarding the use of AI technology in medical translation procedures—a critical effort aligning with Saudi Arabia's Vision 2030 that focuses on technological advancements as well as delivering high-quality healthcare services. By bringing to light the strengths and weaknesses of applying AI technology to translate medical jargon into Arabic language, this study provides a foundation on which to develop better, accurate, and ethically sound applications of AI in the new healthcare environment in the region.

ACKNOWLEDGMENT

This research received grant no. (380/2024) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

REFERENCES

Al-Jarf, R. (2018). Multiple Arabic equivalents to English medical terms: Translation issues. *International Linguistics Research*, *1*(1), 102-110.

Almahasees, Z., & Husienat, I. (2024). A comparative analysis of terminological inconsistency in scientific translation from English into Arabic across different medical fields. *Training, Language and Culture*, *8*(3), 25-40.

Azizova, S. (2023). *Challenges of Medical Translation in the 21st Century from English into Azerbaijani* (Doctoral dissertation).

Banner, D., Bains, M., Carroll, S., Kandola, D. K., Rolfe, D. E., Wong, C., & Graham, I. D. (2019). Patient and public engagement in integrated knowledge translation research: are we there yet?. *Research involvement and engagement*, *5*, 1-14.

Dahal, S. B., & Aoun, M. (2023). Exploring the Role of Machine Translation in Improving Health Information Access for Linguistically Diverse Populations. *Advances in Intelligent Information Systems*, *8*(2), 1-13.

Elhadary, T. (2023). Linguistic and Cultural Differences Between English and Arabic Languages and their Impact on the Translation Process. *International Journal of Language and Translation Research*, *3*(2), 103-117.

Halliday, M.A.K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M.A.K., & Matthiessen, C.M.I.M. (2013). *Halliday's Introduction to Functional Grammar* (4th ed.). Routledge.

Karliner, L. S., Jacobs, E. A., Chen, A. H., & Mutha, S. (2007). Do professional interpreters improve clinical care for patients with limited English proficiency? A systematic review of the literature. *Health services research*, *42*(2), 727-754.

Leonardi, L. (2022). The importance of accurate medical translation in the context of the COVID-19 pandemic. *Revista Granite Journal*, *7*(1), 1-13.

Nasution, A. H., & Onan, A. (2024). Chatgpt label: Comparing the quality of human-generated and llm-generated annotations in low-resource language nlp tasks. *IEEE Access*.

Nazi, Z. A., & Peng, W. (2024, August). Large language models in healthcare and medical domain: A review. *Informatics,11*, 3, 57-68). MDPI.

Olimat, S. N. M. (2019). *Euphemism in the Qur'an: Corpus-based linguistic analysis and intratextual-and contextual-based translation*, Doctoral dissertation, University of Leeds.

Ponnusamy, V., Vasuki, A., Clement, J. C., & Eswaran, P. (2022). AI-Driven Information and Communication Technologies, Services, and Applications for Next-Generation Healthcare System. *Smart Systems for Industrial Applications*, 1-32.

Roba, A. A. (2023). *Challenges in Contextual Translation of Lexical Items from English into Arabic*, Doctoral dissertation, University of Nairobi.

Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, *69*, 343-418.

Sun, G., & Zhou, Y. H. (2023). AI in healthcare: navigating opportunities and challenges in digital communication. *Frontiers in digital health*, *5*, 1291132.

Xiaoqing Mao, A., & Thakkar, I. (2023). Lost in Translation: The Vital Role of Medical Translation in Global Medical Communication. *AMWA Journal: American Medical Writers Association Journal*, *38*(3).

Yaseen, H. S. S. A. (2013). *Terminological inconsistency in medical translation from English into Arabic* (Doctoral dissertation).

Ye, L. (2024). The Feasibility Study of Artificial Intelligence ChatGPT in Translation Field. *Frontiers in Computing and Intelligent Systems*, *8*(1), 52-57.