

تقييم فعالية أدوات الترجمة الصوتية في ترجمة المعلومات الطبية من الإنجليزية إلى العربية

زكريا المحاسيس¹ و اسلام حسينات²

¹ قسم الترجمة، كلية الاداب، جامعة اليرموك، اربد، الأردن

² مستشفى الملك المؤسس عبد الله الجامعي، اربد، الأردن

المخلص

تهدف هذه الدراسة إلى تقييم فعالية أداتي الذكاء الاصطناعي Gemini و DeepSeek في ترجمة المعلومات الطبية الصوتية من الإنجليزية إلى العربية، مع التركيز على دقة نقل المحتوى الطبي وجودة المخرجات اللغوية. اعتمدت الدراسة منهجًا وصفيًا تحليليًا مقارنًا، وتكونت البيانات من 40 مثالًا طبيًا صوتيًا تم اختيارها من محتوى طبي منشور على YouTube، وتضمنت موضوعات متنوعة، منها الأمراض والأعراض، والأدوية والجراحات، والفحوصات، والتحذيرات، والتعليمات الطبية. تم تقديم المقاطع الصوتية إلى النظامين بصورة مستقلة، ثم تحليل المخرجات وفق إطار Multidimensional Quality Metrics (MQM)، مع التركيز على بُعدي Fluency و Accuracy وتصنيف الأخطاء إلى فئات، أبرزها الترجمة غير الدقيقة، والحذف، والإضافة، والمشكلات اللغوية. كما تم اختيار عشرة أمثلة تمثيلية لإجراء تحليل نوعي مقارن أكثر تفصيلًا. أظهرت النتائج قدرة النظامين على إنتاج ترجمات عربية واضحة وطبيعية في معظم الحالات، مع ظهور تفاوت في المحافظة على بعض المصطلحات والتفاصيل الطبية ودرجات الاحتمال والاستعجال. كما بين التحليل أن الطلاقة اللغوية لا تضمن دائمًا الدقة الطبية، إذ يمكن أن تكون الترجمة سليمة لغويًا مع وجود تغير دلالي في المعلومات الأصلية. وتخلص الدراسة إلى أن Gemini و DeepSeek يمثلان أدوات واعدة لدعم الترجمة الطبية، إلا أن استخدامهما في السياقات الحساسة يستلزم مراجعة بشرية متخصصة لضمان دقة المعلومات وسلامتها.

الكلمات المفتاحية

الترجمة الطبية؛ الترجمة الصوتية؛ الذكاء الاصطناعي؛ Gemini؛ DeepSeek؛ الترجمة الآلية؛ الترجمة من الإنجليزية إلى العربية؛ مقياس جودة الترجمة متعدد الأبعاد (MQM)؛ الدقة؛ الطلاقة.

Assessing the Effectiveness of Voice-Translation Tools for Translating Medical Information from English to Arabic

Zakaryia Almahasees¹ and Islam Husienat²

¹ Department of Translation, Yarmouk University, Irbid, Jordan

² King Abdullah University Hospital, Irbid, Jordan

Corresponding Author: Pending confirmation, **E-mail:** zmhases@hotmail.com

ABSTRACT

In this study, the performance of Gemini and DeepSeek is evaluated in translating spoken medical information from English into Arabic. Specifically, special attention is paid to medical and linguistic aspects of translation. A descriptive, analytical, and comparative approach is adopted in examining a corpus of 40 spoken medical excerpts, which was collected from public sources of medical information on YouTube. In general, the corpus covers various areas of medicine including diseases and symptoms, medicines and doses, medical tests, warnings, and instructions. Each spoken medical excerpt was translated into Arabic independently by Gemini and DeepSeek, and the resulting translation outputs were assessed with the help of the MQM approach paying special attention to such criteria as Accuracy and Fluency. Errors in medical translation were classified into several groups, namely mistranslation, omission, addition, and linguistic errors. In order to complement error analysis, ten representative excerpts were chosen and carefully analyzed qualitatively. Results showed that both systems produce clear, fluent, and natural Arabic translations; nevertheless, the systems differ in the ability to preserve medical terms, clinical

information, degree of probability, and urgency. It should be emphasized that the analysis demonstrated that fluency and medical accuracy in translation do not coincide as fluent and natural translations may still be semantically different from the source information. Therefore, it is crucial to make the distinction between linguistic and semantic characteristics of translations produced by AI. The current study proves that Gemini and DeepSeek have great potential for becoming valuable tools for supporting English-Arabic medical translation especially concerning spoken medical information. However, their employment in the sphere of medicine is possible only with the specialized human evaluation.

KEYWORDS

Medical translation; Speech translation; Artificial intelligence; Gemini; DeepSeek; Machine translation; English–Arabic translation; Multidimensional Quality Metrics (MQM); Accuracy; Fluency.

ARTICLE INFORMATION

ACCEPTED: 02 September 2026

PUBLISHED: 27 September 2026

DOI: 10.32996/ijilt.2026.9.9.22

المقدمة

شهدت الترجمة خلال العقدین الماضیین تحولات تقنية متسارعة نتيجة التطور في الذكاء الاصطناعي (Artificial Intelligence)، ومعالجة اللغة الطبيعية (Natural Language Processing)، والتعرف الآلي على الكلام (Automatic Speech Recognition)، والترجمة الآلية العصبية (Neural Machine Translation). وقد أسهمت هذه التطورات في انتقال الترجمة الآلية من الأنظمة القائمة على القواعد والترجمة الإحصائية إلى نماذج أكثر قدرة على معالجة السياق وإنتاج نصوص تتسم بدرجات متقدمة من الطلاقة والترابط. ومع ظهور النماذج اللغوية الكبيرة (Large Language Models)، اتسع نطاق استخدام الذكاء الاصطناعي في الترجمة ليشمل التعامل مع النصوص التخصصية، وتحليل السياق، وإعادة الصياغة، والتلخيص، والتفاعل متعدد اللغات، الأمر الذي أدى إلى تزايد الاهتمام بمدى إمكانية استخدام هذه التقنيات في مجالات تتطلب مستويات مرتفعة من الدقة، ومن أبرزها المجال الطبي (فيلالي، 2026).

وتعد الترجمة الطبية أحد أكثر مجالات الترجمة التخصصية حساسية؛ نظرًا إلى طبيعة المعلومات التي تتعامل معها والعواقب المحتملة للأخطاء في نقلها. فهي لا تقتصر على إيجاد مقابلات لغوية للمصطلحات الطبية، وإنما تتطلب المحافظة الدقيقة على المعلومات المتعلقة بالتشخيص، والأعراض، والأدوية، والجرعات، والتحذيرات، والإجراءات الطبية، والتاريخ المرضي، ونتائج الفحوصات، والتعليمات الموجهة إلى المرضى. ومن ثم، فإن الخطأ في ترجمة المعلومات الطبية قد يتجاوز كونه مشكلة لغوية ليصبح مشكلة تواصلية أو سريرية محتملة، خاصة إذا أدى إلى تغيير جرعة دوائية، أو حذف النفي، أو تغيير درجة الاحتمال، أو تقليل درجة الاستعجال في تعليمات الطوارئ (عميرة & الحسين، 2024).

وتزداد أهمية الترجمة الطبية في البيئات الصحية متعددة اللغات، حيث يمكن أن تشكل الاختلافات اللغوية عائقًا أمام التواصل الفعال بين المرضى ومقدمي الرعاية الصحية. وقد يؤدي عدم فهم المريض للمعلومات المقدمة إليه إلى صعوبة في اتباع التعليمات الطبية أو فهم التشخيص أو الالتزام بالعلاج. وفي هذا السياق، وفرت أدوات الترجمة الرقمية وسيلة سريعة نسبيًا لتقليل بعض الحواجز اللغوية، ولا سيما في الحالات التي لا يتوافر فيها مترجم طبي أو مترجم فوري محترف. إلا أن سهولة الوصول إلى هذه الأدوات لا تعني بالضرورة ملاءمتها لجميع أشكال التواصل الصحي؛ إذ تظل الدقة والموثوقية والسلامة من القضايا المركزية عند تقييم إمكانية استخدامها في المجال الطبي (Ibrahim, 2024).

وقد أثارت هذه المسألة اهتمام الباحثين منذ المراحل المبكرة لانتشار تطبيقات الترجمة الآلية. فقد قام (Sandhu et al., 2022) بتقييم استخدام Google Translate في ترجمة العبارات الطبية، ووجد الباحثان أن نسبة الترجمات التي اعتبرت صحيحة بلغت 57.7% في العينة التي اختبروها، مع وجود تفاوت ملحوظ في الأداء بين اللغات. وقد دفعت هذه النتائج الباحثين إلى التحذير من الاعتماد غير المشروط على الترجمة الآلية في التواصل الطبي. وعلى الرغم من أن هذه الدراسة أجريت قبل التطورات الكبيرة في الترجمة الآلية العصبية والنماذج اللغوية الكبيرة، فإنها أسست لسؤال بحثي ما يزال قائمًا، وهو مدى إمكانية الوثوق بالمخرجات الآلية عندما تكون المعلومات المراد نقلها ذات طبيعة طبية.

ومع التطور اللاحق في الترجمة الآلية، أظهرت الدراسات الحديثة تحسنًا كبيرًا في جودة المخرجات مقارنة بالأجيال السابقة. فعلى سبيل المثال، أظهرت دراسة (Carreras Tartak et al., 2026)، التي تناولت استخدام Google Translate في ترجمة تعليمات الخروج من قسم الطوارئ عبر 26 لغة، مستوى مرتفعًا من النجاح الإجمالي في نقل المعلومات. ومع ذلك، لم تكن الدقة متساوية في جميع اللغات وجميع أنواع التعليمات، وهو ما يشير إلى أن التحسن العام في أداء أنظمة الترجمة لا يلغي الحاجة إلى تقييمها على مستوى اللغة والمجال ونوع النص.

وتصبح هذه القضية أكثر أهمية عند الانتقال من الترجمة النصية إلى الترجمة الصوتية (Speech Translation). فالترجمة الصوتية تتضمن معالجة الكلام المنطوق وتحويله إلى معلومات مفهومة بلغة أخرى، وقد تمر العملية بعدة مراحل تقنية، أهمها التعرف الآلي على الكلام وتحويله إلى نص، ثم ترجمة المحتوى إلى اللغة الهدف، وقد تتبعها مرحلة تحويل النص المترجم إلى كلام. ويعني ذلك أن جودة المخرج النهائي لا تعتمد على كفاءة الترجمة وحدها، وإنما تتأثر كذلك بدقة التعرف على الكلام. فإذا أخطأ النظام، على سبيل المثال، في التعرف على اسم دواء أو قيمة رقمية أو مصطلح طبي في التسجيل الأصلي، فقد تنتقل هذه المشكلة إلى مرحلة الترجمة حتى إذا كانت قدرة النموذج على الترجمة نفسها مرتفعة (يعقوب & الشراح، 2009).

وتشكل هذه الخاصية تحديًا مهمًا في الخطاب الطبي المنطوق، نظرًا إلى كثافة المصطلحات التخصصية وأسماء الأدوية والاختصارات والأرقام ووحدات القياس. كما قد تتأثر أنظمة التعرف على الكلام بعوامل مثل سرعة الحديث، واللهجة، وجودة التسجيل، والضوضاء المحيطة، وطريقة نطق المصطلحات الطبية. ولذلك فإن تقييم فعالية أدوات الترجمة الصوتية في المجال الطبي ينبغي أن يميز منهجيًا بين أخطاء التعرف على الكلام (ASR errors) وأخطاء الترجمة (translation errors)، لأن عدم الفصل بين المرحلتين قد يؤدي إلى نسبة الخطأ إلى مكون غير مسؤول عنه مباشرة (قن & خديجة، 2022).

ويكتسب اتجاه الترجمة من الإنجليزية إلى العربية أهمية خاصة في هذا السياق. فالإنجليزية تمثل لغة رئيسية في إنتاج المعرفة الطبية والبحث العلمي والمواد الصحية الرقمية، بينما تمثل العربية لغة التواصل الأساسية للملايين المرضى في الدول العربية. وفي الوقت نفسه، تختلف اللغتان بصورة واضحة من حيث الأنظمة الصرفية والنحوية والمعجمية، كما أن المصطلحات الطبية العربية لا تتسم دائمًا بدرجة واحدة من التوحيد؛ إذ قد يستخدم المصطلح العربي المتخصص إلى جانب المصطلح الإنجليزي أو المقابل المعرب أو التعبير المبسط الموجه إلى المريض (رفيقة، 2016).

وقد أظهرت الدراسات التي تناولت الترجمة الطبية إلى العربية وجود تحديات مرتبطة بهذا الاتجاه اللغوي. فعلى سبيل المثال، تناول (Delfani et al., 2024) ترجمة معلومات الصحة النفسية من الإنجليزية إلى عدد من اللغات، من بينها العربية، وأظهر التحليل وجود صعوبات مرتبطة بترجمة المصطلحات الطبية والطلاقة وقابلية الفهم. وتشير هذه النتائج إلى أن تقييم النظام ينبغي ألا يعتمد على جودة أدائه العامة في الترجمة، وإنما على قدرته الفعلية على معالجة النص الطبي ضمن الزوج اللغوي المحدد. كما يمثل توفر الموارد المتخصصة عاملاً أساسيًا في تطوير وتقييم الترجمة الطبية الإنجليزية-العربية. وفي هذا الإطار، قدمت (عميرة & الحسين، 2023) (مدونة PEACH، وهي مدونة متوازنة متخصصة في الرعاية الصحية تضم 51,671 زوجًا من الجمل الإنجليزية-العربية. ويوفر وجود مثل هذه الموارد أساسًا مهمًا لبناء مجموعات اختبار أكثر تخصصًا وإجراء تقييمات تستند إلى نصوص طبية حقيقية بدل الاختصار على الجمل العامة أو المصطنعة).

وقد أضاف ظهور النماذج اللغوية الكبيرة بعدًا جديدًا إلى دراسة الترجمة الطبية. فهذه النماذج لا تتعامل مع الترجمة باعتبارها مجرد استبدال لوحدات لغوية بين لغتين، وإنما تستطيع الاستفادة من السياق والتعليمات المقدمة إليها وإنتاج صيغ مختلفة وفقًا للمهمة. ويمكن، على سبيل المثال، توجيه النموذج إلى ترجمة نص طبي إلى العربية الفصحى، أو تبسيطه للمريض، أو المحافظة على المصطلحات المتخصصة. وتوفر هذه المرونة إمكانات مهمة، لكنها تثير في الوقت نفسه إشكاليات تتعلق باحتمال الحذف أو الإضافة أو إعادة الصياغة بما يتجاوز حدود النص (المصدر (عميرة & الحسين، 2023)).

وتؤكد منظمة الصحة العالمية (Organization, 2024)، في إرشاداتها المتعلقة باستخدام النماذج متعددة الوسائط الكبيرة في الصحة، ضرورة إخضاع هذه الأنظمة لمبادئ السلامة والشفافية والحوكمة والمراجعة المناسبة. ويعود ذلك، من بين أسباب أخرى، إلى قدرة النماذج التوليدية على إنتاج استجابات تبدو لغويًا متماسكة ومقنعة في الوقت الذي قد تحتوي فيه على معلومات غير دقيقة. وتكتسب هذه المسألة أهمية خاصة في الترجمة الطبية؛ لأن الطلاقة اللغوية لا تعادل بالضرورة الدقة السريرية.

وفي دراسة حديثة ذات صلة مباشرة بموضوع البحث الحالي، قارن (Li et al., 2026) بين النماذج اللغوية الكبيرة وأدوات الترجمة الآلية التقليدية في ترجمة ملخصات الاستشارات الطبية إلى العربية والصينية والفيتنامية. وأظهرت الدراسة أن جودة الترجمة تتأثر باللغة والأداة ودرجة تعقيد المحتوى، كما أبرزت اختلاف أداء المقاييس الآلية في الحكم على جودة الترجمات. وتؤكد هذه النتائج الحاجة إلى الجمع بين التقييم الآلي والتقييم البشري وتحليل الأخطاء عند دراسة الترجمة الطبية.

وتدعم ذلك النتائج التي توصل إليها (Abouzahir et al., 2026) في تقييمهم متعدد اللغات للنماذج اللغوية في المهام الطبية العربية والإنجليزية؛ إذ كشفت الدراسة عن وجود فجوة مرتبطة باللغة في أداء النماذج، ولا سيما مع زيادة تعقيد المهمة. وتكتسب هذه النتيجة أهمية منهجية للدراسة الحالية؛ لأنها تشير إلى أن الأداء المرتفع للنموذج في المهام الطبية باللغة الإنجليزية لا يمكن اعتباره دليلًا كافيًا على تحقيق المستوى نفسه عند معالجة المحتوى الطبي باللغة العربية.

ومن بين النماذج التي استقطبت اهتمامًا متزايدًا في السنوات الأخيرة Gemini و DeepSeek فقد تطورت عائلة Gemini باتجاه قدرات متعددة الوسائط تشمل معالجة النص والصوت وغيرها من المدخلات، بينما برز DeepSeek بوصفه أحد النماذج اللغوية الحديثة ذات القدرات المتقدمة في الاستدلال ومعالجة النصوص. وقد أدى انتشار النموذجين إلى زيادة استخدامهما في المهام اللغوية والمعرفية المختلفة، بما في ذلك الترجمة ومعالجة المعلومات الطبية. ومع ذلك، فإن إثبات كفاءة النموذج في مهام الاستدلال أو الإجابة عن الأسئلة الطبية لا يعني بالضرورة كفاءته بالدرجة نفسها في الترجمة الطبية؛ إذ تتطلب الأخيرة المحافظة الدقيقة على المحتوى الموجود في النص المصدر وعدم إدخال معلومات جديدة أو حذف معلومات قائمة (Kim & Park, 2024).

وتزداد الحاجة إلى دراسة هذين النموذجين في ضوء محدودية الأبحاث التي تتناول بصورة مباشرة ومقارنة قدرتهما على ترجمة المعلومات الطبية من الإنجليزية إلى العربية. فمع أن الدراسات الحديثة بدأت في تقييم النماذج اللغوية الكبيرة في السياقات الصحية وفي الترجمة الطبية متعددة اللغات، ما تزال الدراسات التي تجمع بين Gemini و DeepSeek، والترجمة الطبية، والاتجاه الإنجليزي-العربي، والمعلومات ذات الأصل الصوتي محدودة. وتمثل هذه المحدودية فجوة بحثية واضحة، خاصة مع الانتشار السريع لاستخدام تطبيقات الذكاء الاصطناعي من قبل غير المتخصصين (فارح et al., 2022).

وعلاوة على ذلك، فإن عددًا من مقاييس تقييم الترجمة الآلية التقليدية يركز على التشابه بين المخرج والترجمة المرجعية، في حين أن هذا النوع من القياس لا يكفي وحده للحكم على السلامة الطبية. فقد يستخدم النموذج مصطلحًا مختلفًا عن الترجمة المرجعية ويظل صحيحًا، بينما يمكن في حالة أخرى أن ينتج جملة شديدة التشابه مع المرجع لكنها تحتوي على خطأ في رقم أو نفي أو جرعة أو درجة احتمال. ولذلك تتطلب الترجمة الطبية منهجًا أكثر شمولًا يجمع بين الدقة الدلالية (semantic accuracy)، والدقة المصطلحية (terminological accuracy)، واكتمال المعلومات (information completeness)، والطلاقة (fluency)، والوضوح (clarity)، والمحافظة على درجة الاستعجال والاحتمالية (clinical urgency and modality).

وانطلاقًا من هذه الفجوة، تهدف الدراسة الحالية إلى تقييم فعالية Gemini و DeepSeek في ترجمة المعلومات الطبية من الإنجليزية إلى العربية من خلال إطار تحليلي متعدد الأبعاد يركز على الدقة الطبية واللغوية والتواصلية. كما تسعى إلى تحديد أنماط الأخطاء التي يمكن أن تظهر في المخرجات، مثل الحذف، والإضافة، والخطأ المصطلحي، والتضيق أو التوسع الدلالي، والاستبدال المفاهيمي، وتغيير درجة الاحتمال، وفقدان درجة الاستعجال في التعليمات الطبية.

ولا تهدف المقارنة إلى تحديد النظام الذي ينتج النص العربي الأكثر طلاقة فحسب، وإنما إلى تحديد مدى قدرة كل نموذج على المحافظة على سلامة المعلومات الطبية أثناء انتقالها من الإنجليزية إلى العربية. وتفترض الدراسة أن هذا النوع من التقييم يقدم صورة أكثر واقعية عن ملاءمة أدوات الذكاء الاصطناعي للترجمة الطبية، كما يمكن أن يساهم في تحديد حدود استخدامها والحالات التي تستلزم المراجعة البشرية المتخصصة. وبذلك تسعى الدراسة إلى الإسهام في الأدبيات المتنامية حول الذكاء الاصطناعي والترجمة الطبية، مع تقديم تركيز خاص على زوج لغوي لا يزال بحاجة إلى مزيد من البحث التجريبي، وهو الإنجليزية-العربية.

الدراسات السابقة

شهدت الدراسات المتعلقة بالترجمة الآلية في المجال الطبي تحولًا ملحوظًا منذ عام 2020، بالتزامن مع التطور المتسارع في الترجمة الآلية العصبية (Neural Machine Translation) وظهور النماذج اللغوية الكبيرة (Large Language Models). ولم يعد الاهتمام البحثي مقتصرًا على قياس مدى التشابه بين الترجمة الآلية والترجمة البشرية، بل امتد إلى قضايا أكثر تعقيدًا، مثل الدقة المصطلحية، وقابلية الفهم، والسلامة السريرية، والفروق بين اللغات، ومدى ملاءمة المخرجات للتواصل مع المرضى. وتكتسب هذه التحولات أهمية خاصة للدراسة الحالية؛ إذ تتطلب ترجمة المعلومات الطبية من الإنجليزية إلى العربية تقييمًا يتجاوز الطلاقة اللغوية إلى تحديد ما إذا كانت الأنظمة تحافظ على المعلومات السريرية ووظيفة الرسالة الطبية.

الترجمة الآلية في السياقات الطبية

قدم (Vieira et al., 2021) مراجعة نقدية لاستخدام الترجمة الآلية في المجالات عالية الخطورة، مع التركيز بصورة خاصة على القطاعين الطبي والقانوني. وانطلقت الدراسة من حقيقة أن سهولة الوصول إلى أنظمة الترجمة الآلية شجعت على استخدامها خارج المجالات اليومية لتدخل بيئات يكون فيها الخطأ ذا عواقب محتملة على الأفراد. وأظهر تحليل الأدبيات أن استخدام الترجمة الآلية في المستشفيات ومؤسسات الرعاية الصحية يثير قضايا تتجاوز جودة اللغة، لأن أخطاء الترجمة يمكن أن تؤثر في قدرة المريض على فهم المعلومات أو اتخاذ القرارات المناسبة. وأكد الباحثون أن تقييم التكنولوجيا ينبغي ألا يقتصر على الدقة التقنية للنظام، وإنما يجب أن يشمل السياق الاجتماعي والمهني الذي تستخدم فيه الترجمة. وتكتسب هذه الدراسة أهمية بالنسبة للبحث الحالي لأنها تدعم النظر إلى الترجمة الطبية بوصفها عملية عالية المخاطر (high-risk translation setting)، حيث لا يمكن تقييم Gemini و DeepSeek على أساس الطلاقة وحدها. ومع ذلك، ركزت دراسة Vieira et al. على الآثار المجتمعية العامة للترجمة الآلية ولم تقدم تقييمًا تجريبيًا مباشرًا للنماذج اللغوية الحديثة أو للاتجاه الإنجليزي-العربي. ومن ثم، توفر الدراسة أساسًا نظريًا لمفهوم المخاطر، بينما تتجه الدراسة الحالية إلى قياس هذه المخاطر على مستوى المخرجات الفعلية وتحليل طبيعة الأخطاء.

وقد وسع (Zappatore and Ruggieri, 2024) هذا الاتجاه من خلال مراجعة منهجية متعددة المعايير لتبني الترجمة الآلية في قطاع الرعاية الصحية. تناول الباحثان الأدبيات المتعلقة باستخدام تقنيات الترجمة الآلية في البيئات الصحية، مع التركيز على منهجيات التقييم ومدى ملاءمة هذه الأنظمة للاستخدام الفعلي. وأظهرت المراجعة أن الترجمة الآلية يمكن أن تقدم حلًا منخفض التكلفة لتقليل الحواجز اللغوية، ولا سيما عند التعامل مع النصوص منخفضة الحساسية. ومع ذلك، خلص الباحثان إلى أن أنظمة الترجمة العصبية الحديثة ما تزال تثير مخاوف تتعلق بالسلامة عند استخدامها في النصوص الطبية المتخصصة. ومن أهم إسهامات الدراسة تأكيد ضرورة تطوير معايير أكثر انتظامًا لتقييم الترجمة الطبية. فقد أظهرت المراجعة وجود تفاوت بين الدراسات في أنواع النصوص المستخدمة، والأزواج اللغوية، ومقاييس الجودة، وطرق التقييم البشري. وتدعم هذه النتيجة التصميم المنهجي للدراسة الحالية، التي تجمع بين الدقة الدلالية، والدقة المصطلحية، واكتمال المعلومات، والطلاقة، والوضوح، وتحليل الأخطاء وتقدير خطورتها الطبية بدل الاعتماد على مقياس واحد للجودة.

وفي السياق نفسه، أجرى (Merx, Phillips, and Suominen, 2024) مراجعة نطاقية (scoping review) حول استخدام تقنيات الترجمة الآلية في المجال الصحي. وشملت مراجعتهم الدراسات المحكمة المنشورة حول استخدام الترجمة الآلية في مجموعة من البيئات الصحية، بما في ذلك التواصل بين المريض ومقدم الرعاية، والتعليم الطبي، وترجمة الأدبيات والمعلومات الصحية. وأظهرت المراجعة أن التحسن في دقة الأنظمة الحديثة أدى إلى إعادة النظر في إمكانية استخدامها في مواقف كانت الترجمة الآلية تعد فيها سابقًا غير مناسبة. إلا أن الباحثين أكدوا استمرار الحاجة إلى الحذر، نظرًا إلى تفاوت الأداء باختلاف اللغة والسياق ونوع المعلومات الطبية. وتشير هذه المراجعات مجتمعة إلى تحول في السؤال البحثي. فلم يعد السؤال الرئيس هو ما إذا كانت الترجمة الآلية تستطيع ترجمة المعلومات الطبية من حيث المبدأ، وإنما متى تكون الترجمة موثوقة، وما أنواع الأخطاء التي تستمر في الظهور، وما مستوى المخاطر المرتبط بهذه الأخطاء؟ وتمثل هذه النقطة انطلاقًا أساسية للدراسة الحالية.

ركز اتجاه آخر من الدراسات الحديثة على ترجمة المعلومات التي تقدم مباشرة إلى المرضى، وهي فئة مهمة لأنها تجمع بين الحاجة إلى الدقة الطبية والحاجة إلى الوضوح وسهولة الفهم. فقد أجرى (Carreras Tartak et al., 2025) دراسة واسعة نسبيًا لتقييم دقة Google Translate في ترجمة تعليمات الخروج من أقسام الطوارئ. وأعد الباحثون تعليمات خروج لثلاث حالات افتراضية، ثم قيموا الترجمات عبر 26 لغة من خلال

مراجعين يتقنون الإنجليزية واللغة الهدف. وشمل التقييم مجالات أساسية من تعليمات الخروج، مثل التشخيص، والأدوية، والمتابعة، والتحذيرات المتعلقة بالأعراض التي تستوجب العودة إلى قسم الطوارئ. أظهرت الدراسة معدل نجاح كمي إجماليًا بلغ 95%، وهو ما يشير إلى تحسن واضح في قدرة أنظمة الترجمة الحديثة على التعامل مع المعلومات الطبية مقارنة بالأجيال السابقة. إلا أن هذه النتيجة العامة أخفت تفاوتًا مهمًا بين اللغات؛ ففي 16 لغة من أصل 26 سجل مجال واحد على الأقل من مجالات التعليمات درجة 90% أو أقل، كما سجلت خمس لغات درجة إجمالية أقل من 90%. وخلص الباحثون إلى أن الأداء المرتفع إجماليًا لا يكفي لتبرير الاستخدام غير المشروط للنظام في السياقات الطبية، وأوصوا بإجراء مزيد من الأبحاث المتعلقة بالسلامة. وتكتسب هذه النتائج أهمية مباشرة للدراسة الحالية من جانبين. أولًا، تؤكد أن متوسط الدقة العام قد يخفي أخطاء مهمة داخل أنواع محددة من المعلومات الطبية. وثانيًا، تبين أن الأداء يتأثر باللغة، الأمر الذي يجعل من الضروري إجراء تقييم مستقل للاتجاه الإنجليزي-العربي بدل تعميم النتائج المستخلصة من لغات أخرى. ومع ذلك، اقتصرت دراسة Carreras Tartak et al. على Google Translate ولم تتناول النماذج التوليدية الحديثة مثل Gemini وDeepSeek، وهو ما يمثل أحد أوجه الاختلاف الرئيسية عن الدراسة الحالية.

الترجمة الآلية للمعلومات الطبية إلى العربية

تمثل اللغة العربية تحديًا خاصًا لأن الأدبيات التجريبية حول الترجمة الطبية إليها أقل نسبيًا من بعض اللغات الأوروبية والآسيوية ذات الموارد الرقمية الأكبر. ومن الدراسات المهمة في معالجة هذا النقص دراسة (Al-Sabbagh, 2024)، التي طورت PEACH: Parallel English-Arabic Corpus for Healthcare، وهي مدونة متوازنة ومتخصصة في النصوص الصحية الإنجليزية والعربية. تتكون PEACH من 51,671 زوجًا من الجمل المتوازنة، بإجمالي يقارب 590,517 كلمة إنجليزية و567,707 كلمة عربية. وتشمل البيانات نشرات معلومات المرضى والمواد التثقيفية الصحية، وقد تمت مواءمة الجمل يدويًا، مما يجعل المدونة موردًا مرجعيًا مهمًا لدراسات الترجمة الطبية الإنجليزية-العربية. وأشارت Al-Sabbagh إلى إمكانية استخدام المدونة في بناء المعاجم ثنائية اللغة، وتكييف النماذج اللغوية للترجمة المتخصصة، وتقييم الترجمة الآلية في الرعاية الصحية، ودراسة مدى وضوح المواد الطبية للمرضى. وعلى الرغم من أن دراسة (Al-Sabbagh, 2024) لا تقارن أداء أدوات ترجمة بعضها، فإن أهميتها للدراسة الحالية منهجية وبيانية؛ إذ تؤكد إمكانية بناء تقييم للترجمة الطبية الإنجليزية-العربية استنادًا إلى مواد صحية متخصصة وترجمات عربية موثوقة بدل الاعتماد على نصوص عامة. كما توفر الدراسة دليلًا على الحاجة إلى موارد لغوية متخصصة عند تقييم النماذج الحديثة في العربية.

وفي سياق أكثر ارتباطًا بتحليل الأخطاء، درس (Delfani et al., 2024) استخدام Google Translate في ترجمة معلومات الصحة النفسية من الإنجليزية إلى خمس لغات، شملت العربية والفارسية والتركية والرومانية والإسبانية. واعتمد الباحثون مجموعتين من النصوص الطبية مأخوذتين من مصادر صحية بريطانية موثوقة، ثم أخضعوا المخرجات لتقييم يدوي من متحدثين أصليين باللغات المستهدفة. وركز التقييم على دقة المصطلحات الطبية، وقابلية الفهم، والطلاقة، والأخطاء التركيبية والدلالية الحرجة. أظهرت النتائج أن ترجمة المصطلحات الطبية مثلت تحديًا ملحوظًا، ولا سيما في العربية والرومانية والفارسية. كما ظهرت مشكلات في الطلاقة أثرت في قابلية الفهم، وكانت هذه المشكلات بارزة في العربية. وخلص الباحثون إلى أن التحسن العام في الترجمة الآلية لا يلغي الحاجة إلى المراجعة البشرية عند ترجمة المحتوى الصحي. وتتصل هذه الدراسة بصورة وثيقة بالبحث الحالي لأنها توضح أن تقييم العربية ينبغي أن يشمل المصطلح الطبي والدقة الدلالية وقابلية الفهم في الوقت نفسه. كما أن استخدامها لتحليل الأخطاء يدعم المنهج النوعي المستخدم في الدراسة الحالية. ومع ذلك، اقتصرت الدراسة على Google Translate ولم تختبر النماذج اللغوية التوليدية الحديثة، كما ركزت على مجال الصحة النفسية، في حين تشمل الدراسة الحالية طيفًا أوسع من المعلومات الطبية.

النماذج اللغوية الكبيرة والترجمة الطبية

أدى انتشار النماذج اللغوية الكبيرة إلى ظهور مرحلة جديدة من الأبحاث في الترجمة الطبية. وتختلف هذه النماذج عن أنظمة الترجمة التقليدية في قدرتها على التعامل مع تعليمات المستخدم والسياق وإعادة الصياغة، وهو ما يوفر إمكانيات إضافية ولكنه يخلق في الوقت نفسه مخاطر مرتبطة بالحذف والإضافة وتغيير المعلومات. ومن أكثر الدراسات ارتباطًا بالدراسة الحالية بحث (Li et al., 2026)، المنشور في JMIR Formative Research، والذي قارن النماذج اللغوية الكبيرة بأنظمة الترجمة الآلية التقليدية في ترجمة ملخصات الاستشارات الطبية من الإنجليزية إلى العربية والصينية والفيتنامية. وشملت الدراسة ثلاثة نماذج لغوية كبيرة هي GPT-4o وGemini-2-7B وLlama-3.1-405B، وثلاثة أنظمة ترجمة تقليدية هي Google Translate وMicrosoft Bing Translator وDeepL.

اعتمد الباحثون على نصوص طبية بدرجات مختلفة من التعقيد، بما يسمح بدراسة أثر طبيعة المحتوى والجمهور المستهدف في جودة الترجمة. وأظهرت النتائج أن أداء الأنظمة لم يكن ثابتًا عبر اللغات أو درجات التعقيد، وأن النماذج اللغوية والأنظمة التقليدية أظهرت نقاط قوة مختلفة بحسب مقياس التقييم المستخدم. كما أبرزت الدراسة محدودية الاعتماد على المقاييس الآلية وحدها في الحكم على جودة الترجمة الطبية. وتعد دراسة (Li et al., 2026) الأقرب من حيث التصميم إلى البحث الحالي، لأنها تتعامل مباشرة مع ترجمة معلومات طبية من الإنجليزية إلى العربية وتقارن تقنيات ذكاء اصطناعي حديثة. ومع ذلك، توجد ثلاثة اختلافات أساسية. أولًا، لم تشمل الدراسة Gemini أو DeepSeek. ثانيًا، كان التركيز على ترجمة النصوص المكتوبة وليس المعلومات ذات الأصل الصوتي. ثالثًا، اعتمدت بصورة رئيسة على تقييم كمي للجدوى، بينما تضيف الدراسة الحالية تحليلًا نوعيًا تفصيليًا للأخطاء، يشمل التحول في درجة الاحتمال، والحذف الدلالي، والتضيق الدلالي، والاستبدال المفاهيمي، وفقدان درجة الاستعجال، والحذف الإجرائي.

أداء النماذج اللغوية في المجال الطبي العربي

لا يكفي تقييم قدرة النموذج على الترجمة العامة لتحديد مدى موثوقيته في العربية الطبية؛ إذ يمكن أن تؤثر خصائص اللغة نفسها في الأداء. وقد عالجت الدراسات الحديثة هذه القضية بصورة مباشرة. قدم (Daoud et al., 2025) معيار MedArabiQ لتقييم النماذج اللغوية الكبيرة في المهام الطبية العربية. ويتكون المعيار من سبع مهام طبية باللغة العربية تغطي تخصصات متعددة، وتشمل أسئلة الاختيار من متعدد، وملء الفراغات، والأسئلة والأجوبة بين المريض والطبيب. وقارن الباحثون سبعة نماذج مفتوحة ومغلقة، من بينها Gemini 1.5، بهدف تحديد قدرة النماذج الحديثة على التعامل مع المحتوى الطبي العربي. وتبرز أهمية هذه الدراسة في أنها لا تفترض أن الأداء المرتفع للنموذج في الإنجليزية

سينتقل تلقائيًا إلى العربية، بل تستخدم معيارًا عربيًا مستقلاً. كما توضح الحاجة إلى موارد ومعايير تقييم مصممة خصيصًا للعربية الطبية. ومع أن MedArabiQ لا يقيس الترجمة، فإنه يوفر دليلًا مهمًا على أن تقييم النماذج في السياق الطبي العربي ينبغي أن يتم بصورة مستقلة. وطور (Abouzahir et al. (2026) هذا الاتجاه من خلال إجراء تقييم تجريبي عابر للغات لأداء النماذج اللغوية في المهام الطبية باللغتين العربية والإنجليزية. وأظهرت الدراسة وجود فجوة أداء مستمرة مرتبطة باللغة، وأن هذه الفجوة تصبح أكثر وضوحًا كلما ازدادت المهمة تعقيدًا. كما كشف تحليل الترميز (tokenization) عن درجة أكبر من التجزئة البنيوية في النص الطبي العربي، بينما أظهر تحليل الموثوقية أن الثقة التي يعبر عنها النموذج وتفسيراته ليست مؤشرات موثوقة دائمًا على صحة المخرج. وتعد هذه النتيجة مهمة بصورة خاصة للترجمة الطبية؛ إذ يمكن أن ينتج النموذج نصًا عربيًا طليقًا وواثقًا ظاهريًا رغم احتوائه على تغيير دلالي أو مصطلحي. ولذلك ينبغي ألا يستخدم fluency مؤشرًا بديلًا عن accuracy. وتتبنى الدراسة الحالية هذا التمييز من خلال تقييم الطلاقة والدقة الدلالية والمصطلحية بوصفها متغيرات منفصلة.

التركيب النقدي للدراسات السابقة

تكشف مراجعة الدراسات الحديثة عن تطور واضح في النظرة إلى استخدام الترجمة الآلية والذكاء الاصطناعي في الرعاية الصحية. فقد أظهرت الدراسات والمراجعات الحديثة أن التقنيات الحالية أكثر قدرة من الأنظمة السابقة على إنتاج ترجمات قابلة للفهم، إلا أن الأدلة لا تدعم الاستخدام غير المشروط لها في جميع السياقات الطبية. ويظهر اتفاق عام بين الدراسات على أن نوع اللغة، ونوع النص، ودرجة التعقيد، وطبيعة المعلومات الطبية، وطريقة التقييم جميعها عوامل تؤثر في الحكم على جودة النظام. كما تكشف الأدبيات عن انتقال منهجي من تقييم الجودة العامة إلى تحليل المخاطر. فقد أظهرت (Carreras Tartak et al. (2025 أن ارتفاع معدل النجاح الإجمالي لا يعني خلو الترجمات من مشكلات داخل أنواع محددة من التعليمات، بينما بينت (Delfani et al. (2024 أن العربية تواجه تحديات خاصة في المصطلحات الطبية والطلاقة. ومن ناحية أخرى، أظهر (Li et al. (2026 أن النماذج اللغوية الحديثة لا تتفوق بالضرورة على أنظمة الترجمة التقليدية في جميع المقاييس واللغات، وأن تقييم الترجمة الطبية يتطلب منهجًا متعدد الأبعاد. وتعزز الدراسات الخاصة بالعربية هذه الصورة. فقد وفرت (Al-Sabbagh (2024 موردًا متخصصًا واسع النطاق للترجمة الصحية الإنجليزية-العربية، بينما أثبت (Abu Daoud et al. (2025 و (Abouzahir et al. (2026 أن تقييم النماذج في المجال الطبي العربي يحتاج إلى معايير مستقلة، وأن أداء النماذج يمكن أن يتأثر بالخصائص اللغوية للعربية.

على الرغم من هذا التقدم، تكشف الدراسات السابقة عن فجوة واضحة يمكن تحديدها في أربعة جوانب مترابطة. أولًا، ما تزال الدراسات التي تتناول الترجمة الطبية من الإنجليزية إلى العربية أقل من الدراسات التي تتناول اللغات ذات الموارد المرتفعة. ثانيًا، تركز نسبة كبيرة من الدراسات على أنظمة الترجمة الآلية التقليدية، بينما لا تزال المقارنات المباشرة بين النماذج التوليدية الحديثة محدودة. ثالثًا، لم تقدم الدراسات التي تمت مراجعتها مقارنة مباشرة بين Gemini و DeepSeek في ترجمة المعلومات الطبية الإنجليزية-العربية. رابعًا، تركز غالبية الدراسات على النصوص المكتوبة، في حين لم تحظ المعلومات الطبية ذات الأصل الصوتي بالقدر نفسه من التحليل، ولا سيما فيما يتعلق بالفصل بين أخطاء التعرف على الكلام وأخطاء الترجمة.

المنهجية

اعتمدت الدراسة منهجًا وصفيًا تحليليًا مقارنةً لتقييم فعالية Gemini و DeepSeek في ترجمة المعلومات الطبية الصوتية من الإنجليزية إلى العربية. وتكونت عينة الدراسة من 40 مقطعًا صوتيًا طبيًا باللغة الإنجليزية تم اختيارها قصديًا من مقاطع فيديو طبية منشورة على YouTube. وروعي في اختيار المقاطع تنوع المحتوى الطبي واشتماله على مصطلحات ومعلومات ذات أهمية سريرية، مثل الأمراض والأعراض، والأدوية والجرعات، والفحوصات، والتعليمات الطبية، والتحذيرات، والمعلومات المتعلقة بالطوارئ.

جمع البيانات وإجراءات الترجمة

تم استخراج المقاطع الطبية المختارة من محتوى YouTube والاستماع إليها والتحقق من النص الإنجليزي المنطوق قبل استخدامها في الاختبار. بعد ذلك، تم تقديم المحتوى الصوتي نفسه إلى كل من Gemini و DeepSeek بصورة مستقلة، وطلب من النظامين ترجمة المعلومات المنطوقة من الإنجليزية إلى العربية. وتم جمع أول مخرج ترجمي ينتجه كل نظام دون إجراء أي تعديل أو تصحيح بشري، مما أسفر عن 80 مخرجًا مترجمًا، بواقع 40 ترجمة لكل نظام. ثم تمت مقارنة المخرجات بالنصوص الإنجليزية المصدر وبالترجمات العربية المرجعية بهدف تحديد الاختلافات والأخطاء وتحليلها.

منهجية تحليل الأخطاء

اعتمد تحليل المخرجات على إطار Multidimensional Quality Metrics (MQM) (Lommel et al., 2015) في تصنيف أخطاء الترجمة وتحديد طبيعتها. وركزت الدراسة بصورة أساسية على بُعدين رئيسيين هما Accuracy و Fluency، لارتباطهما المباشر بدقة نقل المعلومات الطبية وجودة النص العربي الناتج.

يشمل بُعد Accuracy الأخطاء التي تؤثر في العلاقة بين النص المصدر والترجمة، مثل Omission، Addition، و Untranslated Text. وأولي اهتمام خاص للأخطاء التي تؤدي إلى تغيير المصطلحات الطبية، أو الأرقام والجرعات، أو النفي، أو درجة الاحتمال، أو التحذيرات والتعليمات الطبية. أما بُعد Fluency فيركز على جودة اللغة العربية الناتجة بصرف النظر عن النص المصدر، ويشمل مشكلات Grammar، Word Order، Spelling، Punctuation، و Register/ Stylistic appropriateness عندما تؤثر في سلامة المخرج ووضوحه.

التقييم	فئات الأخطاء المستخدمة في التحليل	محور التقييم
Accuracy	Mistranslation, Omission, Addition, Untranslated Text	مدى المحافظة على المعنى والمعلومات والمصطلحات الطبية في النص المصدر
Fluency	Grammar, Word Order, Spelling, Punctuation, Register/Style	سلامة اللغة العربية وطبيعية الصياغة ووضوح المخرج

تم تحديد الأخطاء في كل مخرج وترميزها وفق الفئة المناسبة، ثم حساب عدد الأخطاء وتكراراتها ونسبها المئوية في كل من Gemini و DeepSeek. وبعد ذلك، أجريت مقارنة بين النظامين لتحديد أكثر أنواع الأخطاء شيوعاً ومدى اختلاف أدائهما في بُعدي Accuracy و Fluency. ولتعزيز التحليل النوعي، تم اختيار 10 أمثلة تمثيلية من أصل 40 مثالاً لمناقشتها بالتفصيل. وعُرض كل مثال في جدول أفقي يشتمل على النص الإنجليزي المصدر، والترجمة العربية المرجعية، وترجمة Gemini، وترجمة DeepSeek، وتصنيف الخطأ وفق MQM، ثم أعقب كل جدول تحليل يوضح طبيعة الخطأ وأثره في الدقة الطبية وجودة الترجمة. وقد أتاح الجمع بين التكرارات الكمية والتحليل النوعي تقديم تقييم أكثر شمولاً لأداء النظامين في ترجمة المعلومات الطبية الصوتية من الإنجليزية إلى العربية.

التحليل والمناقشة

اعتمد التحليل على إطار Multidimensional Quality Metrics (MQM)، مع التركيز على بُعدي Accuracy و Fluency. وتمت مقارنة كل ترجمة بالنص الإنجليزي المصدر لتحديد مواضع الخطأ وتصنيفها، مع إيلاء اهتمام خاص لمدى المحافظة على المصطلحات والمعلومات الطبية. وفيما يلي تحليل تفصيلي للأمثلة المختارة.

المثال الأول: احتشاء عضلة القلب الحاد (Acute Myocardial Infarction (AMI)

النص المصدر (Source Text)	Gemini	DeepSeek
The patient was diagnosed with acute myocardial infarction.	شخص المريض باحتشاء عضلة القلب الحاد.	شخص المريض بنوبة قلبية حادة.

يكشف المثال الأول اختلافاً مصطلحياً في ترجمة acute myocardial infarction. فقد قدم Gemini المقابل "احتشاء عضلة القلب الحاد"، وهو مقابل طبي مباشر يحافظ على المكونات الدلالية للمصطلح الإنجليزي؛ إذ ينقل myocardial إلى "عضلة القلب"، و infarction إلى "احتشاء"، و acute إلى "الحاد". ومن منظور Accuracy في إطار MQM، لا يظهر في هذا المخرج تغيير جوهري في المعلومات الطبية. أما DeepSeek فقد استخدم "نوبة قلبية حادة"، وهي صياغة شائعة ومفهومة في الخطاب الصحي العام، إلا أنها أقل تخصصاً من المصطلح الطبي الوارد في المصدر. ويمكن النظر إلى هذا الاختيار باعتباره تحولاً من المصطلح الطبي التقني إلى تعبير أكثر عمومية. ومع ذلك، لا يصل الاختلاف بالضرورة إلى مستوى Mistranslation الكامل، لأن "النوبة القلبية" تستخدم للإشارة إلى الحالة نفسها في الخطاب غير المتخصص.

أما من حيث Fluency، فقد أنتج النظامان جملتين سليميتين نحويًا وطبيعيتين في العربية، ولم تظهر مشكلات واضحة في التركيب أو ترتيب الكلمات أو الإملاء. ومن ثم، يكشف المثال عن أهمية عدم الخلط بين الطلاقة والدقة المصطلحية؛ فقد تكون الترجمة سليمة وطبيعية لغويًا مع وجود اختلاف في مستوى التخصص الطبي. ويبدو Gemini في هذا المثال أكثر محافظة على terminological specificity للنص المصدر، بينما يقدم DeepSeek صياغة أكثر تبسيطًا. ويكتسب هذا الفرق أهمية أكبر إذا كان النص جزءًا من تقرير أو سجل طبي متخصص، حيث يفضل الحفاظ على المصطلح الطبي الدقيق.

المثال الثاني: الجرعة الدوائية وتكرار الاستخدام Medication Dosage and Dosing Frequency

النص المصدر (Source Text)	Gemini	DeepSeek
Take 5 mg twice daily after meals.	تناول جرعة 5 ملغ مرتين يوميًا بعد الطعام.	تناول 5 ملغ مرتين يوميًا بعد الوجبات.

يختبر هذا المثال قدرة النظامين على المحافظة على ثلاثة عناصر مهمة في التعليمات الدوائية: الجرعة (5 mg)، والتكرار (twice daily)، والتوقيت (after meals). وقد حافظ Gemini على العناصر الثلاثة دون حذف أو إضافة مؤثرة، كما نقل DeepSeek الجرعة والتكرار والتوقيت بصورة صحيحة. وإضافة كلمة "جرعة" في مخرج DeepSeek لا تغير المحتوى الموجود في النص المصدر، بل تجعل العبارة أكثر وضوحاً في العربية. وبناءً على تصنيف MQM، لا يظهر في المخرجين خطأ من فئات Mistranslation أو Omission أو Addition يؤثر في Accuracy؛ إذ بقيت المعلومات الدوائية الأساسية محفوظة بالكامل.

وفيما يتعلق بـ Fluency، تتميز الترجمتان بسلامة التركيب ووضوح التعليمات. ويعد الاختلاف بين "بعد الوجبات" و "بعد الطعام" اختلافاً معجمياً مقبولاً ولا يؤثر في الوظيفة الإرشادية للنص. ويشير المثال إلى ضرورة التعامل بحذر مع مفهوم الخطأ في تقييم الترجمة الآلية؛ فالاختلاف عن صياغة مرجعية أو عن مخرج نظام آخر لا يعني تلقائياً وجود خطأ. كما يوضح أهمية المحافظة الدقيقة على القيم العددية في النصوص الطبية، لأن أي تغيير في 5 mg أو twice daily كان سيصنف باعتباره خطأ Accuracy ذا أهمية أكبر بكثير من الاختلاف الأسلوبى الظاهر هنا.

المثال الثالث: التاريخ المرضي لداء السكري Medical History of Diabetes Mellitus

النص المصدر (Source Text)	Gemini	DeepSeek
The patient has no history of diabetes mellitus.	ليس لدى المريض تاريخ مرضي للإصابة بداء السكري.	المريض ليس مصابًا بداء السكري.

يظهر الفرق الرئيس في ترجمة العبارة no history of diabetes mellitus. Gemini حافظ على مفهوم history باستخدام "تاريخ مرضي"، وبذلك نقل المعلومة المتعلقة بالسجل الطبي للمريض بصورة واضحة. أما DeepSeek فقد أنتج "المريض ليس مصابًا بداء السكري"، وهي جملة تشير إلى الحالة الراهنة للمريض ولا تنقل بصورة كاملة فكرة عدم وجود تاريخ مرضي سابق للسكري. ووفق إطار MQM، يمكن تصنيف هذا التحول ضمن Accuracy – Omission/Mistranslation، لأن جزءًا من المحتوى الدلالي المرتبط بالتاريخ المرضي لم يظهر في الترجمة.

ومن حيث Fluency، لا توجد مشكلة لغوية واضحة في أي من المخرجين؛ فجملة DeepSeek سليمة وطبيعية تمامًا باللغة العربية رغم أنها أقل دقة دلاليًا. ويعد هذا المثال مهمًا لأنه يوضح أن الخطأ في الترجمة الطبية قد يكون غير ظاهر للقارئ عند تقييم النص الهدف وحده. فإذا تم تقييم المخرج وفق الطلاقة فقط، فقد يحصل DeepSeek على تقييم مرتفع، في حين تكشف المقارنة بالنص المصدر عن فقدان معلومة طبية. ومن ثم يدعم المثال ضرورة الفصل في MQM بين Accuracy و Fluency وعدم استخدام جودة الصياغة العربية بوصفها دليلًا كافيًا على صحة نقل المعلومات الطبية.

المثال الرابع: الآثار الجانبية ودرجة الاحتمال Adverse Drug Effects and Clinical Modality

النص المصدر (Source Text)	Gemini	DeepSeek
This medication may cause dizziness and drowsiness.	قد يسبب هذا الدواء الدوار والنعاس.	يسبب هذا الدواء الدوار والنعاس.

يتركز التحليل في هذا المثال على ترجمة may cause. Gemini حافظ على أداة الاحتمال من خلال "قد يسبب"، وبالتالي احتفظ بدرجة الاحتمالية الموجودة في المصدر. أما DeepSeek فقد حذف المقابل الدلالي لـ may وأنتج "يسبب هذا الدواء"، وهو ما يحول العلاقة من احتمال حدوث الأثر الجانبي إلى تقرير أكثر قطعية. ويمكن تصنيف هذا الاختلاف ضمن Accuracy – Mistranslation وفق MQM، لأن المعلومة الأساسية لم تحذف بالكامل، ولكن القوة الدلالية للعبارة تغيرت. ويكتسب ذلك أهمية في النص الطبي لأن التمييز بين may cause و causes يعكس اختلافًا حقيقيًا في درجة اليقين.

أما فيما يتعلق بـ Fluency، فإن المخرجين سليمين لغويًا ولا توجد مشكلة في النحو أو ترتيب الكلمات أو الإملاء. والمفارقة هنا أن حذف عنصر دلالي واحد لم يجعل ترجمة DeepSeek أقل طلاقة؛ بل بقيت الجملة طبيعية تمامًا باللغة العربية. ولذلك يقدم المثال دليلًا تطبيقيًا على أن أخطاء Accuracy في الترجمة الطبية قد توجد داخل نصوص تتمتع بمستوى مرتفع من Fluency. ويشير ذلك إلى ضرورة فحص modality بعناية عند تقييم ترجمة التحذيرات والآثار الجانبية للأدوية.

المثال الخامس: إيقاف الدواء دون استشارة الطبيب Medication Discontinuation and Treatment Adherence

النص المصدر (Source Text)	Gemini	DeepSeek
Do not discontinue the medication without consulting your physician.	لا تتوقف عن تناول الدواء دون استشارة طبيبك.	لا توقف الدواء قبل التحدث مع الطبيب.

يحافظ كلا النظامين على الوظيفة الأساسية للنص، وهي منع المريض من إيقاف العلاج دون الرجوع إلى الطبيب. وينقل Gemini discontinue the medication إلى "تتوقف عن تناول الدواء"، بينما يستخدم DeepSeek "توقف الدواء". كما يقابل Gemini consulting your physician بـ "استشارة طبيبك"، في حين يستخدم DeepSeek "التحدث مع الطبيب". وعلى الرغم من اختلاف الصياغة، لا يظهر حذف جوهرية للمعلومة أو تغيير للإجراء المطلوب. ولذلك لا يستدعي الاختلاف في هذا المثال تصنيفًا واضحًا ضمن أخطاء Accuracy في MQM.

أما على مستوى Fluency، فإن مخرج Gemini أكثر اصطلاحية قليلًا في سياق التعليمات الطبية بسبب استخدام "استشارة طبيبك"، إلا أن صياغة DeepSeek تظل واضحة ومفهومة. ويبرهن المثال على ضرورة ألا يتحول تطبيق MQM إلى عملية بحث ميكانيكية عن الاختلافات بين المخرجات؛ إذ ينبغي التمييز بين الاختلاف المقبول (acceptable variation) والخطأ الفعلي. فإذا حافظت الصياغة البديلة على المعنى الطبي والإجراء المطلوب ولم تتضمن حذفًا أو إضافة أو ترجمة خاطئة، فإن الاختلاف الأسلوبى وحده لا ينبغي احتسابه خطأ في Accuracy.

المثال السادس: القصور الكلوي وتعديل الجرعة Renal Impairment and Dose Adjustment

النص المصدر (Source Text)	Gemini	DeepSeek
Patients with renal impairment may require dose adjustment.	قد يحتاج المرضى الذين يعانون من قصور كلوي إلى تعديل الجرعة.	قد يحتاج مرضى الفشل الكلوي إلى تغيير الجرعة.

يتمثل الاختلاف الرئيس في ترجمة المصطلح renal impairment. Gemini قصور كلوي"، وهو مقابل يحافظ بدرجة أكبر على المعنى العام للمصطلح الذي يشير إلى وجود ضعف أو اختلال في وظائف الكلى بدرجات متفاوتة. أما DeepSeek فقد استخدم "الفشل الكلوي"، وهو تعبير أكثر تحديدًا وشدة من المصطلح الأصلي. ومن منظور **Accuracy** في إطار MQM، يمكن تصنيف هذا التحول ضمن **Mistranslation** أو التضييق الدلالي، لأن الترجمة انتقلت من مفهوم واسع نسبيًا إلى حالة مرضية أكثر شدة. كذلك حافظ النظامان على عنصر الاحتمال may require، ولم يظهر حذف في هذا الجزء من الجملة.

أما من حيث **Fluency**، فقد جاءت الترجمتان سليمتين وطبيعتين في العربية، ولم تظهر مشكلات في البنية النحوية أو ترتيب الكلمات. غير أن سلامة اللغة في مخرج DeepSeek لا تلغي الإشكال المصطلحي، وهو ما يؤكد مرة أخرى أن الطلاقة لا تعني بالضرورة الدقة الطبية. ويكتسب هذا المثال أهمية خاصة لأن درجة القصور الكلوي قد تكون مرتبطة بتعديل الجرعات الدوائية، ولذلك فإن استخدام مصطلح أكثر شدة من الأصل قد يؤثر في تفسير الحالة الطبية.

المثال السابع: صعوبة التنفس والحاجة إلى رعاية طبية عاجلةDyspnea and Urgent Medical Attention

النص المصدر (Source Text)	Gemini	DeepSeek
Seek immediate medical attention if you experience difficulty breathing.	اطلب العناية الطبية فورًا إذا واجهت صعوبة في التنفس.	راجع الطبيب إذا شعرت بصعوبة في التنفس.

يحافظ Gemini على المكونات الأساسية في النص المصدر، ولا سيما عبارة immediate medical attention التي نقلها إلى "العناية الطبية فورًا". أما DeepSeek فقد استخدم "راجع الطبيب"، وهي صياغة تنقل الحاجة إلى استشارة طبية لكنها لا تحافظ على درجة الاستعجال الموجودة في المصدر. ووفق إطار MQM، يمكن تصنيف هذا التحول ضمن **Accuracy – Omission/Mistranslation**، لأن عنصر immediate لم يُنقل بصورة واضحة، كما أن طبيعة الإجراء تحولت من طلب رعاية فورية إلى مراجعة عامة للطبيب.

وعلى مستوى **Fluency**، فإن الترجمتين واضحتان وسليمتان لغويًا، غير أن مخرج DeepSeek يغير القوة التداولية للتعليمات الطبية. وتظهر أهمية ذلك في أن النصوص المتعلقة بالأعراض الحادة لا تنقل معلومات فقط، بل توجه المتلقي إلى سلوك محدد وفي توقيت محدد. ولذلك فإن فقدان درجة الاستعجال، رغم عدم تأثيره في الطلاقة، يمثل مشكلة مهمة في الترجمة الطبية ويمكن أن يكون أكثر خطورة من العديد من الأخطاء الأسلوبية.

المثال الثامن: ارتفاع إنزيمات الكبدElevated Liver Enzymes

النص المصدر (Source Text)	Gemini	DeepSeek
The blood test showed elevated liver enzymes.	أظهر فحص الدم ارتفاع مستويات إنزيمات الكبد.	أظهر تحليل الدم زيادة في وظائف الكبد.

يظهر الفرق الأساسي في ترجمة liver enzymes. Gemini استخدم "إنزيمات الكبد"، وهو مقابل مباشر يحافظ على المفهوم الطبي المحدد. أما DeepSeek فقد استخدم "وظائف الكبد"، وهو مفهوم مختلف وإن كان مرتبطًا بالمجال نفسه. ومن منظور **Accuracy** في إطار MQM، يمكن تصنيف هذا الاختلاف ضمن **Mistranslation**، لأن الترجمة استبدلت مفهومًا طبيًا محددًا بمفهوم أوسع لا يعادل النص المصدر. واختبارات وظائف الكبد قد تتضمن مجموعة من المؤشرات، في حين أن النص يشير تحديدًا إلى ارتفاع الإنزيمات.

أما من حيث **Fluency**، فكلتا المخرجات مقبول لغويًا ولا توجد مشكلة في التركيب أو الوضوح. وتكمن خطورة مخرج DeepSeek في أن الخطأ قد لا يكون واضحًا للقارئ غير المتخصص، نظرًا إلى الترابط الدلالي بين "إنزيمات الكبد" و"وظائف الكبد". وبين المثال أن الأخطاء الطبية قد تكون خفية داخل نصوص سليمة لغويًا، وهو ما يدعم الحاجة إلى تقييم الترجمة من قبل مختصين قادرين على التمييز بين المصطلحات المتقاربة طبيًا.

المثال التاسع: الصيام قبل الإجراء الطبيPre-Procedural Fasting

النص المصدر (Source Text)	Gemini	DeepSeek
The patient should fast for eight hours before the procedure.	يجب على المريض الصيام لمدة ثماني ساعات قبل الإجراء.	ينبغي للمريض عدم تناول الطعام لمدة ثماني ساعات قبل الفحص.

يحافظ Gemini على مفهومي fast وprocedure باستخدام "الصيام" و"الإجراء"، كما يحافظ على المدة الزمنية البالغة ثماني ساعات. أما DeepSeek فيفسر fast بعبارة "عدم تناول الطعام" ويترجم procedure إلى "الفحص". ويمكن تصنيف هذا التحول ضمن **Accuracy – Mistranslation**، لأن "الفحص" أضيق دلاليًا من "الإجراء"، الذي قد يشمل إجراءً تشخيصيًا أو علاجيًا أو تدخلًا طبيًا آخر. كما أن مفهوم الصيام الطبي قد لا يقتصر في جميع الحالات على الطعام فقط، ولذلك فإن التفسير قد يؤدي إلى تضييق المعلومات الأصلية.

ومن حيث **Fluency**، فإن مخرج DeepSeek واضح وطبيعي لغويًا، وهو ما قد يجعله مقنعًا للمتلقي رغم احتوائه على تغير دلالي. ويظهر المثال مرة أخرى أهمية المقارنة بالنص المصدر عند تطبيق MQM؛ إذ إن تقييم النص العربي وحده لن يكشف المشكلة. كما يوضح أن التعليمات

التحضيرية للإجراءات الطبية تتطلب درجة عالية من المحافظة على التفاصيل، لأن أي تغيير في نوع الإجراء أو طبيعة الصيام قد يؤثر في فهم المتلقي للتعليمات.

المثال العاشر: ألم الصدر الشديد والاستجابة للطوارئ Severe Chest Pain and Emergency Response

النص المصدر (Source Text)	Gemini	DeepSeek
If you develop severe chest pain, call emergency services immediately.	إذا شعرت بألم شديد في الصدر، فاتصل بخدمات الطوارئ فوراً.	إذا أصبت بألم شديد في الصدر، اطلب المساعدة الطبية.

يحافظ Gemini على العناصر الثلاثة الرئيسية في النص: العرض severe chest pain، والإجراء call emergency services، ودرجة الاستعجال immediately. أما DeepSeek فيحافظ على العرض ويشير بصورة عامة إلى طلب المساعدة الطبية، لكنه لا ينقل بشكل صريح ضرورة الاتصال بخدمات الطوارئ ولا عنصر "فوراً". ووفق تصنيف MQM، يمكن اعتبار ذلك Accuracy – Omission، لأن جزءاً من التعليمات الإجرائية والزمنية لم يظهر في المخرج، كما يمكن اعتباره جزئياً Mistranslation بسبب تحويل الإجراء المحدد إلى نصيحة عامة.

أما فيما يتعلق بـ **Fluency**، فإن ترجمة DeepSeek سليمة ومفهومة من الناحية اللغوية، لكنها أقل دقة وظيفياً من ترجمة Gemini. وتبرز أهمية المثال لأن التعليمات الطبية في حالات الطوارئ لا تقتصر على وصف الحالة، بل تحدد أيضاً الإجراء المطلوب والسرعة اللازمة لتنفيذه. ولذلك فإن حذف وسيلة التصرف أو درجة الاستعجال يمثل خطأ في الدقة قد تكون له أهمية سريرية حتى وإن بقيت الجملة العربية واضحة وطبيعية.

المناقشة

تكشف نتائج التحليل المقارن لمخرجات Gemini و DeepSeek أن تقييم فعالية أدوات الذكاء الاصطناعي في ترجمة المعلومات الطبية من الإنجليزية إلى العربية لا ينبغي أن يعتمد على سلامة النص الهدف أو طلاقته وحدهما، وإنما على مدى المحافظة الدقيقة على المعلومات الطبية الواردة في النص المصدر. وقد أظهر تحليل الأمثلة وفق إطار Multidimensional Quality Metrics (MQM) أن الفروق الأكثر أهمية بين النظامين ظهرت بصورة أساسية في بُعد Accuracy، ولا سيما في ترجمة المصطلحات الطبية، والمحافظة على درجة الاحتمال، والتعليمات الطبية، ودرجة الاستعجال. في المقابل، كانت مشكلات Fluency أقل وضوحاً في الأمثلة النوعية المختارة؛ إذ أنتج النظامان في معظم الحالات تراكيب عربية سليمة ومفهومة، حتى عندما احتوت بعض المخرجات على تغير في المعنى الطبي. وتبرز هذه النتيجة أهمية الفصل بين الدقة والطلاقة عند تقييم الترجمة الطبية المولدة بالذكاء الاصطناعي.

ومن أبرز الأنماط التي ظهرت في التحليل مشكلة الدقة المصطلحية. ففي مثال acute myocardial infarction يحافظ Gemini على المصطلح الطبي المتخصص "احتشاء عضلة القلب الحاد"، بينما استخدم DeepSeek نوبة قلبية حادة، وهي صياغة مفهومة ولكنها أقل تخصصاً. وكان الاختلاف أكثر أهمية في ترجمة renal impairment؛ إذ استخدم DeepSeek الفشل الكلوي، في حين أن المصطلح المصدر يشير بصورة أوسع إلى ضعف أو اختلال وظائف الكلى. وظهر نمط مشابه في ترجمة liver enzymes إلى "وظائف الكبد". وتوضح هذه الحالات أن النظام قد ينتج تعبيراً ينتمي إلى المجال الطبي الصحيح ولكنه لا يحافظ بالضرورة على الحدود المفاهيمية الدقيقة للمصطلح الأصلي. وهذه مسألة مهمة في الترجمة الطبية، حيث قد يكون الفرق بين مصطلحين متقاربين لغوياً ذا أهمية سريرية.

كما أظهر التحليل أن بعض الأخطاء ارتبطت بحذف أو تغيير مكونات دلالية صغيرة ولكنها ذات أهمية كبيرة. ويتضح ذلك في ترجمة الجملة This medication may cause dizziness and drowsiness. فقد حافظ Gemini على may باستخدام "قد"، بينما حذفها DeepSeek، فتحوّلت إمكانية حدوث الأعراض إلى صياغة أكثر قطعية. ووفق MQM، يمثل ذلك مشكلة في Accuracy/Mistranslation حتى وإن لم يؤثر في Fluency. وينطبق الأمر نفسه على مثال التاريخ المرضي للسكري؛ إذ أدى تحويل no history of diabetes mellitus إلى "المريض ليس مصاباً بداء السكري" إلى فقدان مفهوم التاريخ المرضي. وتؤكد هذه الحالات أن الكلمات الوظيفية والعلاقات الدلالية، مثل النفي والاحتمال والزمن، تستحق اهتماماً ماثلاً للمصطلحات الطبية عند تقييم المخرجات.

وكانت المحافظة على الاستعجال السريري والتعليمات الإجرائية من أكثر الجوانب حساسية التي كشفت عنها الأمثلة. ففي جملة Seek immediate medical attention if you experience difficulty breathing، أدى مخرج DeepSeek راجع الطبيب إلى تخفيف درجة الاستعجال الموجودة في المصدر. كما ظهر نمط أكثر وضوحاً في مثال ألم الصدر الشديد، حيث تحولت التعليمات المحددة call emergency services immediately إلى "اطلب المساعدة الطبية". وعلى الرغم من أن الترجمة الأخيرة تبدو منطقية وسليمة لغوياً، فإنها تحذف عنصري أساسيين: الجهة المطلوب الاتصال بها وضرورة التصرف الفوري. وتدل هذه النتيجة على أن تقييم الترجمة الطبية ينبغي أن يأخذ في الاعتبار ليس فقط ما يقوله النص، بل أيضاً ما يطلب من المتلقي فعله.

وفي المقابل، لم تكن جميع الاختلافات بين النظامين أخطاء. ففي مثال الجرعة Take 5 mg twice daily after meals، حافظ النظامان على الجرعة والتكرار والتوقيت، رغم الاختلاف الطفيف بين "بعد الوجبات" و"بعد الطعام". وكذلك في مثال عدم إيقاف الدواء، نقل كلا النظامين الوظيفة الأساسية للتحذير رغم اختلاف الصياغة. وتبرز هذه النتائج أهمية تطبيق MQM بطريقة لا تعامل كل اختلاف لغوي عن الصياغة المرجعية باعتباره خطأ. فالترجمة تسمح بوجود أكثر من صياغة صحيحة، وينبغي أن يكون معيار الحكم الرئيس هو المحافظة على المحتوى الطبي والوظيفة التواصلية وليس التطابق اللفظي.

وتنسجم هذه الملاحظات مع الاتجاه العام في الدراسات الحديثة حول استخدام الترجمة الآلية والذكاء الاصطناعي في السياقات الصحية. فقد بينت مراجعات حديثة أن التحسن الكبير في أنظمة الترجمة لا يلغي مخاطر استخدامها في السياقات الطبية، وأن الجودة تتأثر بنوع النص واللغة والسياق السريري. كما أظهرت الدراسات التي قيمت ترجمة المعلومات الموجهة إلى المرضى أن الأداء الإجمالي المرتفع يمكن أن يتعايش مع أخطاء محددة ذات أهمية في بعض اللغات أو أنواع التعليمات. وهذا يؤكد أن الاعتماد على معدل دقة عام لا يقدم بالضرورة صورة كاملة عن سلامة النظام.

وتكتسب هذه المسألة أهمية أكبر في الترجمة إلى العربية. فقد أشارت دراسات حديثة تناولت معالجة المحتوى الطبي العربي إلى أن أداء النماذج لا يكون متساوياً بالضرورة بين الإنجليزية والعربية، وأن المهام الطبية العربية قد تطرح تحديات إضافية أمام النماذج اللغوية. كما أظهرت الأبحاث المتعلقة بالترجمة الطبية إلى العربية وجود صعوبات في المصطلحات وقابلية الفهم. وتدعم نتائج الدراسة الحالية هذا الاتجاه؛ إذ ظهرت بعض المشكلات في اختيار المقابل العربي الدقيق رغم أن النص الناتج ظل واضحاً وسليماً لغوياً.

ومن النتائج المهمة أيضاً أن Fluency قد تخفي أخطاء Accuracy. فقد كانت معظم المخرجات التي تضمنت مشكلات دلالية أو مصطلحية سليمة من الناحية النحوية وطبيعية في العربية. فعلى سبيل المثال، "أظهر تحليل الدم زيادة في وظائف الكبد" جملة عربية سليمة تماماً، إلا أنها لا تعادل elevated liver enzymes. وكذلك فإن "المريض ليس مصاباً بداء السكري" جملة طبيعية، لكنها لا تنقل no history of diabetes mellitus بصورة كاملة. وتبرز هذه النتيجة تحدياً جوهرياً في استخدام النماذج اللغوية الكبيرة في الترجمة الطبية؛ فارتفاع جودة الصياغة قد يجعل اكتشاف الخطأ أكثر صعوبة بالنسبة للمستخدم غير المتخصص.

وتشير النتائج كذلك إلى أهمية طبيعة البيانات الصوتية المستخدمة في الدراسة. فالترجمة الصوتية تتطلب من النظام معالجة المحتوى المنطوق قبل إنتاج الترجمة، ولذلك يمكن أن يكون الخطأ ناتجاً عن التعرف غير الصحيح على الكلام أو عن عملية الترجمة نفسها. وتزداد هذه المشكلة في المحتوى الطبي الذي يحتوي على مصطلحات متخصصة وأسماء أدوية وأرقام وجرعات. ومن ثم، فإن النتائج ينبغي تفسيرها باعتبارها تقييماً للآداء النهائي للأداة في التعامل مع المعلومات الطبية الصوتية، وليس تقييماً مجرداً للترجمة النصية فقط. ويشكل هذا الجانب إضافة مهمة إلى الدراسات التي اقتصرت على إدخال النصوص المكتوبة إلى أنظمة الترجمة.

ومن منظور مقارن، تشير الأمثلة النوعية المختارة إلى أن Gemini أظهر محافظة أكبر على بعض التفاصيل الطبية والمصطلحية في الحالات التي تمت مناقشتها، بينما ظهرت في بعض مخرجات DeepSeek حالات من التعميم أو التضيق الدلالي أو حذف بعض مكونات التعليمات. ومع ذلك، ينبغي عدم تعميم هذه الملاحظة النوعية وحدها على الأداء الكلي للنظامين؛ إذ يجب أن يستند الحكم النهائي على النظام الأفضل إلى نتائج تحليل جميع الأمثلة الأربعين، بما في ذلك العدد الكلي لأخطاء Fluency و Accuracy وتوزيع فئات Omission and Mistranslation وغيرها. وبذلك تؤدي الأمثلة العشرة وظيفة تفسيرية للأنماط التي تظهر في البيانات الكمية، ولا تحل محل المقارنة الإحصائية الشاملة.

وتؤكد النتائج في مجملها الحاجة إلى المراجعة البشرية المتخصصة (human expert review) عند استخدام أدوات الذكاء الاصطناعي في ترجمة المعلومات الطبية الحساسة. فالنظام الذي ينتج نصاً عربياً طليقاً قد يغير مصطلحاً أو يحذف احتمالاً أو يقلل من درجة استعجال تعليمات طبية دون أن يبدو المخرج غير طبيعي. ولذلك ينبغي النظر إلى Gemini و DeepSeek بوصفهما أدوات يمكن أن تساعد في عملية الترجمة، وليس باعتبار مخرجاتهما بديلاً تلقائياً عن التحقق البشري في السياقات التي يمكن أن تؤثر فيها الترجمة في فهم المريض أو سلامته.

وأخيراً، تبرز الدراسة أهمية استخدام إطار MQM في تقييم الترجمة الطبية؛ إذ أتاح الفصل بين Fluency و Accuracy تحديد حالات لا يمكن اكتشافها من خلال التقييم العام للجودة. كما سمح تصنيف الأخطاء إلى Omission و Mistranslation و Addition وغيرها بتفسير طبيعة نقاط الضعف بدل الاكتفاء بالحكم بأن ترجمة ما "صحيحة" أو "غير صحيحة". وتشير النتائج إلى أن تقييم أدوات الترجمة الطبية المستقبلية ينبغي أن يجمع بين التحليل الكمي للأخطاء والتحليل النوعي للأمثلة، مع إعطاء وزن خاص للأخطاء التي تمس المصطلحات الطبية والجرعات والنفي والاحتمال والتحذيرات وتعليمات الطوارئ. ومن شأن هذا النهج أن يقدم تقييماً أكثر واقعية لفعالية النماذج اللغوية الكبيرة في الترجمة الطبية الإنجليزية-العربية.

الخاتمة

هدفت هذه الدراسة إلى تقييم فعالية Gemini و DeepSeek في ترجمة المعلومات الطبية الصوتية من الإنجليزية إلى العربية، من خلال تحليل 40 مثالاً طبياً مستخرجاً من محتوى طبي منشور على YouTube. واعتمدت الدراسة إطار Multidimensional Quality Metrics (MQM) لتحليل المخرجات وتصنيف الأخطاء، مع التركيز على بُعدي Fluency و Accuracy. وأظهر التحليل أن كلا النظامين قادر على إنتاج ترجمات عربية واضحة وطبيعية في معظم الحالات، إلا أن الطلاقة اللغوية لم تكن دائماً مؤشراً على الدقة الطبية. فقد ظهرت بعض الأخطاء المرتبطة بالترجمة غير الدقيقة، والحذف، واختيار المصطلحات الطبية، والمحافظة على درجة الاحتمال والاستعجال والتعليمات الإجرائية. كما أوضح التحليل النوعي للأمثلة المختارة أن بعض الاختلافات التي تبدو محدودة لغوياً يمكن أن تؤدي إلى تغير مهم في المعنى الطبي أو الوظيفة التواصلية للنص.

وتخلص الدراسة إلى أن أدوات الذكاء الاصطناعي التوليدي، مثل Gemini و DeepSeek، تمتلك إمكانات واعده لدعم الترجمة الطبية الصوتية الإنجليزية-العربية، إلا أن استخدامها في السياقات الطبية الحساسة يتطلب مراجعة بشرية متخصصة، ولا سيما عندما تتضمن المعلومات جرعات دوائية، أو مصطلحات تشخيصية، أو نفيًا، أو احتمالية، أو تحذيرات وتعليمات طارئة. كما تؤكد النتائج أهمية الفصل بين Fluency و Accuracy عند تقييم جودة الترجمة، لأن النص العربي قد يكون سليماً وطليقاً مع احتوائه على خطأ دلالي أو طبي. وتوصي الدراسة بإجراء أبحاث مستقبلية على عينات صوتية أكبر وأكثر تنوعاً، تشمل تخصصات ولهجات وظروف تسجيل مختلفة، مع دراسة أخطاء التعرف الآلي على الكلام بصورة مستقلة عن أخطاء الترجمة للوصول إلى تقييم أكثر شمولاً لموثوقية هذه الأدوات في التواصل الطبي متعدد اللغات.

الدعم البحثي:

حصل هذا البحث على المنحة رقم (330) من المرصد العربي للترجمة التابع لمنظمة الألكسو، وبدعم من هيئة الأدب والنشر والترجمة بالملكة العربية السعودية.

المراجع الانجليزية

- [1] Al-Sabbagh, R. (2024). PEACH: a sentence-aligned Parallel English-Arabic Corpus for Healthcare. *Corpora*, 19(Scott), 395-410.
- [2] Abouzahir, C., Khan, M., Ali-Hassan, H., Ma, C., Saleh, K., Sadqi, Y., Mallat, J., Al-Eisawi, W., Habash, N., & Shamout, F. E. (2026). Bridging the English-Arabic Medical Knowledge Gap: Targeted Low-Rank Adaptation via Causal Layer Selection. *arXiv preprint arXiv:2608.00207*.

- [3] Abouzahir, C., Ma, C., Habash, N., & Shamout, F. E. (2026, March). Cross-Lingual Empirical Evaluation of Large Language Models for Arabic Medical Tasks. In *Proceedings of the 1st Workshop on Linguistic Analysis for Health (HeaLing 2026)* (pp. 158-171).
- [4] Carreras Tartak, J. A., Brewster, R. C., Arango Isaza, D., Berumen Martinez, A., Grafals, A., Adusumilli, P., Fitzgerald, T., Orcutt, R., Nathanson, L. A., & Haimovich, A. D. (2026). Evaluating Spanish Translations of Emergency Department Discharge Instructions by a Large Language Model: Tool Validation and Reliability Study. *JMIR Formative Research*, 10, e79676.
- [5] Delfani, J., Orasan, C., Saadany, H., Temizoz, O., Taylor-Stilgoe, E., Kanojia, D., Braun, S., & Schouten, B. (2024). Google translate error analysis for mental healthcare information: Evaluating accuracy, comprehensibility, and implications for multilingual healthcare communication. *arXiv preprint arXiv:2402.04023*.
- [6] Daoud, M. A., Abouzahir, C., Kharouf, L., Al-Eisawi, W., Habash, N., & Shamout, F. E. (2025). Medarabiq: Benchmarking large language models on arabic medical tasks. *arXiv preprint arXiv:2505.03427*.
- [7] Ibrahim, A. S. A. O. (2024). ثقافة الإعاقة ودور الأنثروبولوجيا في فهم ودراسة أمراض الإعاقة. مجلة البحث العلمي في الآداب, 25(4), 119-173.
- [8] Kim, J., & Park, H. (2024). Adaptive latent diffusion model for 3d medical image to image translation: Multi-modal magnetic resonance imaging study. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV),
- [9] Li, A., Zhou, W., Hoda, R., Bain, C., & Poon, P. (2026). Comparing Large Language Models and Traditional Machine Translation Tools for Translating Medical Consultation Summaries: Quantitative Pilot Feasibility Study. *JMIR Formative Research*, 10(1), e85169.
- [10] Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional quality metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*, (12), 0455-463.
- [11] Organization, W. H. (2024). تصنيف التدخلات والخدمات والتطبيقات الرقمية في مجال الصحة: لغة مشتركة لوصف استخدامات التكنولوجيا الرقمية من أجل الصحة. الطبعة الثانية. World Health Organization.
- [12] Sandhu, S. K., Kouvroukoglou, N., Campodonico, M., & Mehendale, F. (2022). Lost in translation: a caution against using automated and commercial translation services for medical communications. *BMJ Global Health*, 7(1), e008067.
- [13] Scott, M. (2008). 3. Comparing corpora and identifying key words, collocations, and frequency distributions through the Word Smith Tools suite of computer programs. In *Small corpus studies and ELT* (pp. 47-70). John Benjamins Publishing Company.
- [14] Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2021). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 24(11), 1515-1532.
- [15] Human Translation in the Age of (فيلالي، ف). 2026. (الترجمة البشرية في عصر الذكاء الاصطناعي التوليدي: تحولات ورهانات Generative AI: Transformations and Existential Implications. *Journal of Translation & Multilingualism*, 4(01), 1-31.
- [16] قن، & خديجة). 2022. (تجربة ترجمة النصوص الطبية من الإنجليزية إلى العربية في العالم العربي من أين وإلى أين؟. افاق للعلوم, 17(1), 479-488.
- [17] يعقوب، م. & الشراح). 2009. (تجربة المركز العربي للوثائق والمطبوعات الصحية في الترجمة من اللغة الإنجليزية إلى العربية "العلوم الطبية". اللغة العربية, 11(1), 387-396.
- [18] Zappatore, M., & Ruggieri, G. (2024). Adopting machine translation in the healthcare sector: A methodological multi-criteria review. *Computer Speech & Language*, 84, 101582.
- [19] Merx, R., Phillips, C., & Suominen, H. (2024, September). Machine translation technology in health: A scoping review. In *Health. Innovation. Community: It Starts With Us: Papers from the 28th Australian Digital Health and Health Informatics Conference (HIC 2024)–Brisbane, Australia, 5–7 August 2024* (pp. 78-83). 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP: SAGE Publications Pvt. Ltd.

المراجع العربية

- رفيقة، م). 2016. (الترجمة الطبية: مفهوم، واقع و تصور [20] رسالة. ماجستير جامعة. وهران
- عميرة، أ. & الحسين). 2023. (المصطلحات الطبية الفارسية الواردة بكتاب أناتومي لخسرو إبراهيم) الجهاز الهضمي أنموذجًا. (المجلة المصرية للبحوث والدراسات الإسلامية, 10(10), 325-356.
- عميرة، أ. & الحسين). 2024. (أهم المصطلحات الطبية الفارسية الواردة بكتاب أناتومي لخسرو إبراهيم) أعضاء جسم الانسان أنموذجًا(دراسة تأصيلية وصفية كما وردت في المعاجم العربية. (المجلة المصرية للبحوث والدراسات الإسلامية, 11(11), 379-407.
- فارج، إبراهيم، بكوش، & الصالح، م). 2022. (ترجمة المصطلح الطبي من الإنجليزية إلى العربية. الصوتيات [23], 18(2), 17-200-9.