

Peer Review File

Article information: <https://dx.doi.org/10.21037/atm-24-162>

Reviewer A

It addresses a crucial gap in the literature regarding the use of AI in real-time patient-provider communication. The paper is well-structured, informative, and follows PRISMA guidelines.

While I believe the paper is strong overall, I suggest the following revisions to improve clarity and depth:

1. The review includes only 9 studies out of 1,095 identified. While I understand the inclusion criteria were strict, the limited scope raises concerns about generalizability. Further justification for this limited scope would be beneficial.
 - a. **Reply 1:** Thank you for your thoughtful comment. The limited number of studies meeting our inclusion criteria reflects the emerging nature of the field and the currently limited evidence available on AI for clinical translation. However, we aimed to enhance applicability by including studies from multiple countries and across various medical specialties, ensuring a broader perspective despite the limited scope.
 - b. **Changes to the text:** We added a section explaining the exclusion criteria to the methods section (*2.3 Selection of Reports and Study Eligibility*). We also addressed this component in the discussion (*4.7 Strengths and Limitations of this Review*). The page and line numbers given reflect "All Markup" as visible on Microsoft Word (see Page 4, Lines 158-161; Page 11, Lines 462-465).
2. A deeper discussion on the ethical and legal implications of using AI for medical translation is needed, especially regarding patient consent and liability. (Not required)
 - a. **Reply 2:** Thank you for highlighting the importance of ethical and legal considerations in AI-based medical translation. In response, we have added a section in the Discussion addressing these aspects, including current legislative frameworks and ethical concerns issues related to the topic.
 - b. **Changes to the text:** We have incorporated an "*Ethical and Legal Implications*" section (4.5) in the discussion, which examines patient consent, liability, and other key issues associated with using AI translation tools in medical interpretation (see Pages 8-9, Lines 340-352).
3. A quantitative meta-analysis was not performed. While I understand the variability in study designs, an explanation of why it was not feasible would be helpful.

- a. **Reply 3:** Thank you for your comment. Due to the significant heterogeneity in study designs, methodologies, and reported outcomes, performing a quantitative meta-analysis was not feasible. Instead, we adopted a narrative synthesis approach, which allowed us to thoroughly explore, interpret, and contextualize the findings of each study. This approach ensured that the complexity and diversity of the included studies were adequately addressed, providing a comprehensive overview of the current evidence.
- b. **Changes to the text:** We added this explanation in the Methods section under "*Data Synthesis and Analysis*." (see Page 5, Lines 198-201).

Grammar and Spelling Errors:

- Line 6: "lease" should be "Please" in the peer-review section.
- Line 120: "Research units included" should be "The research units included."
 - a. **Reply 4:** Thank you for identifying this error. We have revised the sentence as suggested
 - b. **Changes to the text:** The text now reads, "The research units included..." (see Page 4, Line 121)
- Line 166: "retrieved them thoroughly" should be revised to "thoroughly reviewed them."
 - a. **Reply 5:** Thank you for pointing this out. We have revised the phrase as suggested.
 - b. **Changes to the text:** The text now reads, "thoroughly reviewed them" (see Page 5, Line 171).
- Line 212: "presented" should be "obtained."
 - a. **Reply 6:** Thank you for highlighting this. We have revised the text.
 - b. **Changes to text:** The text has changed to "The data is presented in Table 1." (see Page 6, Line 231)
- Line 326: "a translation an" should be "a translation."
 - a. **Reply 7:** Thank you for your comment. However, upon thorough review of the manuscript, we were unable to locate this specific phrase. It is possible that the text has already been revised. If you need further clarification, we would be happy to address it.
 - b. **Changes to text:** No changes were made as we could not locate the specified text.

In conclusion, I recommend moving forward with minor revisions. The paper is valuable to the field, and addressing these points will enhance its impact.

Reviewer B

1. The authors should shift the focus toward discussing how evolving AI models and mobile platforms will enhance real-time medical translation. This could include predictions on improved accuracy, faster processing times, integration with electronic health records, and potential for real-time speech-to-text-to-translation technologies. A more robust discussion on the future potential of AI, rather than just focusing on past research, would add significant value to the study.

- a. **Reply 1:** Thank you for the great feedback. We agree that focusing on the future potential of AI in real-time medical translation is an excellent suggestion. In response, we have strengthened the discussion section overall, particularly the "*Future Directions*" subsection, to highlight advancements in AI models and mobile platforms. This expanded discussion explores how emerging technologies could improve healthcare accessibility, efficiency, and patient safety.
- b. **Changes to the text:** We revised and expanded the discussion section, particularly under "*Future Directions*," to explore how advancements in AI and mobile platforms could enhance real-time medical translation. The page and line numbers given reflect "All markup" as visible. (see Page 11, Lines 468-480)

2. The current review touches on accuracy, but it doesn't dive deeply into the ability of AI models to explain and interpret complex medical terminology, which is a critical component of medical translations.

- a. **Reply 2:** Thank you for raising this important point. The limited discussion on this topic reflects the emerging nature of the field and the current lack of extensive research specifically addressing the ability of AI models to handle medical terminology. However, we agree that this is a critical area for consideration and have added additional insights by incorporating the available research outside of what was identified in the review. This expanded discussion highlights specific challenges, such as errors in word order and literal translations, and emphasizes the need for further investigation.
- c. **Changes to the text:** We highlighted that our review identified challenges with speech synthesis accuracy due to the presence of medical terminology, and provided additional insights in the discussion under "*Practical Applications in Clinical Settings*" (see Page 8, Lines 335-338)

3. The paper should focus more on how AI can be applied in real-time clinical practice, especially in high-stakes situations like emergency departments, critical care units, and

psychiatric care. The authors should propose practical use cases for real-time AI translation, such as during surgical consultations, in multidisciplinary teams with international staff, or in rural areas where professional interpreters are scarce. They should address the potential of AI to bridge these gaps in real-time, particularly focusing on the handling of urgent, complex, and technical medical communications.

- a) **Reply 3:** Thank you for highlighting this important point. While this paper demonstrates the significant potential of AI in medical translation in the coming years, its current limitations suggest it is not yet ready for widespread application. To address this, we significantly strengthened the Discussion section, providing a deeper exploration of AI's promise across diverse medical settings. Additionally, we expanded our recommendations to include specific scenarios where AI could deliver acceptable translations in the absence of human interpreters, offering practical insights into its role as a last-resort option in resource-limited settings.
- b) **Changes to the text:** We expanded the "*Future Directions*" subsection to provide more specific insights into the potential advancements of AI in medical translation. Additionally, we added a new section titled "*Practical Applications in Clinical Settings*," which highlights AI usage across varied medical environments. (see Page 8, Line 323-339; Page 11, Lines 465-480, 478-480)

4. Many of the included studies are theoretical or based on simulated scenarios. Try to incorporate more real-world case studies or examples of successful AI translation tool implementation. This would provide a more practical perspective and show how the findings of the review translate to clinical practice.

- a) **Reply 4:** Thank you for this feedback. We agree that highlighting real-world applications is important to demonstrate the current utility of AI in medical translation. In response, we expanded the "*Practical Applications in Clinical Settings*" section to include examples of successful AI translation tool implementations
- b) **Changes to the text:** We added real-world case examples of AI translation tool implementations to the "*Practical Applications in Clinical Settings*" section. (see Page 8, Lines 323-328).

5. Speech recognition systems can be used not only for interpretation but also for timeouts in the operating room. Discuss related applications related to the following paper: "<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7144990/>"

- a) **Reply 5:** Thank you for this suggestion. We recognize the importance of highlighting additional applications of speech recognition systems. In response, we cited the suggested paper and expanded the discussion to emphasize the broader potential of speech recognition systems in critical

healthcare settings, particularly their role in enabling hands-free language interpretation and enhancing efficiency.

- b. **Changes to the text:** We incorporated the suggested reference and added the following to the discussion. (see Page 11, Lines 472-476).

Reviewer C

The authors present a very nice overview of the use of AI tools in medical translations at point of care. The topic is relevant due to the increase in multi-cultural societies with language barriers, as is pointed out by the authors. I do have a few minor remarks that should be addressed:

1. In the methods it is stated that for google scholar there were too many hits and only the first 100 were selected, this is very arbitrary. A better explanation is needed as this does not lead to reproducible results as google scholar might have a different sorting for another researcher.

- a. **Reply 1:** Thank you for your feedback. We appreciate the opportunity to clarify this point. Given the large number of results retrieved from Google Scholar (15,800 in total), we reviewed and imported only the first 100 results to ensure feasibility while maintaining relevance. Studies beyond this point were observed to significantly decrease in relevance to the research topic. To enhance transparency and reproducibility, we used the default relevance-based sorting method provided by Google Scholar at the time of the search, ensuring consistency in our approach.
- b. **Changes to text:** We added details under the "Databases Searched" section in the Methods to clarify the rationale for selecting the first 100 results from Google Scholar and to specify the use of the default relevance-based sorting method. The page and line number reflect "All markup" as visible in Microsoft Word. (see Page 4, Lines 146-149).

2. The results are all presented under the header "discussion" this is confusing. It would fit better to present results under the results section and end with the conclusion.

- a. **Reply 2:** To improve the clarity and structure of the manuscript, we have reorganized the content by moving the information originally presented in the discussion to the results section. Additionally, we strengthened the discussion section to focus on current AI language translation platforms, practical applications and limitations, and ethical and legal considerations.
- b. **Changes to the text:** The results have been relocated to align with standard manuscript structure. The discussion section has been revised (see Pages 6-12, Lines 232-497).

3. With the rise of large language models (e.g Chatgpt, claude, gemini) these tools are also relatively often used for translations. I understand that this is beyond the scope of the review,

but a short mention or brief discussion on this topic would greatly benefit the manuscript.

- a. **Reply 3:** We appreciate the suggestion to include a brief discussion on the role of large language models (LLMs) in translation. In response, we added a new section titled "*Potential of Large Language Models in Translation*" to the discussion. This section highlights recent studies on the strengths and limitations of LLMs, such as ChatGPT, in translation, with a focus on their applicability to healthcare and the need for broader language coverage. We also expanded upon the translation platforms evaluated in our review.
- b. **Changes to the text:** We added two new sections in the discussion: *Translation Platforms Evaluated* and *Potential of Large Language Models in Translation*. (see Page 8, Lines 304-321).

Reviewer D

This research systematically addresses the language barriers AI models face in translating medical documentation accurately and efficiently, aiming to prevent misunderstandings and harm to patients. The issue of language barriers in healthcare access is a crucial topic. The methodology follows the PRISMA reporting checklist. However, more detail should be provided regarding the tools used to evaluate the studies included. The authors assess the accuracy, acceptability, and usability of AI-driven language translation in clinical settings. However, the evaluation was based on a limited number of studies, with the Google Scholar search result truncated to 100, suggesting that more studies could potentially be included in the analysis.

In more detail, I have several concerns

1-The accuracy and usability scores should be outlined in the Method section and detailed in the supplemental information. Other tools used such as ROBINS-I, the JBI Critical Appraisal Checklist, or MMAT should also be included in the supplemental materials. The author should explain why 2 studies were not assessed using ROBIN-I.

- a. **Reply 1:** Thank you for your feedback. In response, we have added definitions and details for the accuracy and usability scores to the Methods section under "*Data Synthesis and Analysis*." Additionally, we provided an explanation of the bias assessment tools used in the "*Risk of Bias Assessment*" section. To enhance transparency, these tools have been included in the supplemental materials.
- b. **Changes to the text:** We added the definitions and scoring details as outlined above. The Methods section also includes an explanation for why two studies were not assessed using ROBINS-I. The page and line numbers reflect "All markup" as visible on Microsoft Word. (see Page 5, Lines 178-181, 188-192).

2-The sentence "information relevant to our research question" as stated by the authors, needs

further clarification and detailed explanation.

- a. **Reply 2:** Thank you for pointing this out. To address this, we provided a detailed definition and clarification of what this constitutes in the methods section under "*Data Synthesis and Analysis*."
- b. **Changes to the text:** We clarified the phrase "information relevant to our research question" in the methods section. (see Page 5, Line 188)

3-The results section is too short. It should be extended to cover key areas such as accuracy, differences between theoretical scenarios and real-world scenarios, the clinical context, limitations, and ethical considerations. A narrative summary of the main results presented in Table 1 should also be provided.

- a. **Reply 3:** Thank you for this valuable feedback. We agree that a more comprehensive presentation of the results is essential. To address this, we provided a narrative summary of the main findings, including key results from Table 1. Furthermore, we expanded the discussion section to delve deeper into the clinical context, limitations, and ethical considerations, ensuring a thorough and nuanced analysis of these critical aspects.
- b. **Changes to the text:** The results section now includes a narrative summary of the main findings. Additionally, the discussion section was revised as stated. (see Pages 5-7, Lines 232-289; Pages 8-9, Lines 303-497)

4-Differences between translators should be discussed and included in the supplemental information.

- a. **Reply 4:** Thank you for this suggestion. To address this, we added two sections to the Discussion: "*Translation Platforms Evaluated*" and "*Potential of Large Language Models in Translation*." These sections provide a detailed discussion of the differences between translation platforms, highlighting their strengths and limitations.
- b. **Changes to the text:** We expanded the discussion by adding two new sections as detailed above. (see Page 8, Line 303-321)

5-The authors mention that the ethics and legal barriers were not addressed; however, these issues should be analyzed, particularly in studies involving patients.

- a. **Reply 5:** Thank you for your important comment. We agree that ethical and legal considerations are critical to the responsible implementation of AI translation tools in healthcare. In response, we added a dedicated section titled "*Ethical and Legal Implications*" to the Discussion. This section makes note of key issues, including patient confidentiality, data protection, transparency, informed consent, and accountability. We also discuss existing and emerging regulatory frameworks aimed at addressing these challenges.

- b. **Changes to the text:** A new section, "*Ethical and Legal Implications*," was added to the discussion, which examines legal and ethical considerations related to AI translation tools. (see Page 8-9, Line 340-352).

6-The section on Future directions should be expanded. Which standardized metric do the authors recommend? What system should be used to evaluate the level of understanding? How do structural limitations, such as those in emergency rooms or pediatric settings, impact the results?

- a. **Reply 6:** We appreciate this thoughtful feedback and see value in adding specific measures to address these important points. In response, we expanded the "*Future Directions*" section to explore the potential for AI translation tools to provide real-time communication across diverse clinical scenarios, including emergency rooms and pediatric settings. We also specified scenarios in which AI translation may be most effective, recommended a standardized metric for evaluating translation performance, and proposed a system for assessing understanding.
- c. **Changes to the text:** The future directions section has been augmented with the specifics highlighted above. (see Pages 11-12, Lines 469-497).

7-Figure 1 and Figure 6 are not particularly useful and should be deleted.

- a. **Reply 7:** Thank you for your feedback regarding the figures. We believe that both Figure 1 and Figure 6 add value to the manuscript by enhancing reader engagement and providing important visual context. Figure 1 visually depicts how AI translation can be applied in healthcare scenarios, offering a clear and accessible representation of its potential use. Figure 6 demonstrates the critical components that must be considered when analyzing AI translation tools, which we believe is essential for understanding the complexities of evaluating these systems. While these figures do not present key results, they serve to complement the text and make the paper more engaging and informative for readers.
- b. **Changes to the text:** No changes were made.

8-Figure 2 is already included in the text and does not need to be repeated.

- a. **Reply 8:** Thank you for pointing this out. We agree that including Figure 2 is redundant, as the information is already provided in the text. In response, Figure 2 may be removed from the manuscript.
- b. **Changes in the text:** Deletion of Figure 2 legend and renumbering of subsequent figures.