Is Artificial Intelligence Ready to Tackle Language Barriers for Health Economics and Outcomes Research?

Yeh D[1], Roeder C[2], Longo R[2], Chamoux C[2], Manuel F[1], Saito K[3]
[1]AESARA Inc., Chapel Hill, NC, USA, [2]AESARA Europe, London, UK, [3]University of North Carolina, Chapel Hill, NC, USA.

CONTACT US

BACKGROUND

The health economics and outcomes research (HEOR) community is a global community of researchers and professionals who have various linguistic and cultural backgrounds. Although English is the predominant language for HEOR publications, language can still pose a barrier for researchers. Additionally, the terminology utilized in HEOR research contains many technical words that are not used in everyday language. Artificial intelligence (AI) has evolved significantly, impacting many areas, including translation services. Despite the rapid development of AI, researchers have highlighted the importance of engaging specialists in the specific scientific domain and specialized language.[1]

OBJECTIVE

This study aims to assess the accuracy of AI translation in HEOR publications.

METHOD

Five English-language abstracts from published manuscripts covering various HEOR disciplines were chosen:

- Cost Effectiveness Analysis (CEA) - Ophthalmology[2]
- Real World Evidence (RWE) - Respiratory[3]
- Systematic Literature Review (SLR) - COVID-19[4]
- Randomized Controlled Trial (RCT) - Rare diseases[5]
- Patient Experience Data (PED) - Gastrointestinal disease[6]

Each abstract was translated into Chinese, French, German, Italian, and Japanese using OpenAI's ChatGPT 3.5 in May 2024

Five HEOR researchers, who were native speakers of these languages, assessed the quality of translations. The review focused on HEOR jargon and whether the translations preserved the original meanings in English. Reviewers also ranked the quality of translated abstracts by type of research

The translated abstracts were also translated back to English (reverse translation) to identify discrepancies with the original abstract

Translation Languages

Chinese · French · German · Italian · Japanese

RESULTS

Overall assessment

Reviewers' rankings of abstract translation quality varied by type of research, which could be explained by the language structure and the amount of research literature available in the non-English language (see Figure 1).

Level of comprehension

Overall, reviewers were able to comprehend the main takeaways of the abstracts across all languages except French, where key HEOR terms were often translated inaccurately.

Research Jargon

Although the methods and results sections were generally translated properly, research jargon including "follow-up", "efficacy", "treatment naïve", "washout", "dominant", and "quality-adjusted life years" was not always translated to reflect the original meaning in English, nor maintained given that those terms are usually translated from English.

Reverse Translation results

After reverse translation, the abstracts were very consistent with the original ones with minor missing and inaccurate errors. In Japanese and Chinese, disease and drug names were inconsistently translated to reflect the common terms in those languages.

Presented at ISPOR Europe 2024 Conference, November 17-20, Barcelona, Spain

Figure 1: Translation Assessment by Study Type

	Chinese	French	German	Italian	Japanese
Best Translated Article Type	PRO, RCT	N/A	SLR	SLR	PRO, RCT
Worst Translated Article Type	RWE	SLR, CEA	RWE	CEA	CEA

Table 1: Assessment by Language

Language	Review Summary
Chinese	• Overall, the translation was done properly, the abstracts kept the original meaning in most situation • In the methodology section, AI appropriately used common Chinese words in the methodology section. Technical jargons are consistent with the ones used in Chinese. • For the results portion, if it was written in long sentence form, sentences were translated awkwardly. Sometimes the true meaning may be missing due to the language structure. • Some words like "follow up" 随访, "subject" 主体 are translated awkwardly in Chinese • For the RWE abstract, the same drug name (dupilumab) was used in 3 different words, mixing Japanese or Chinese characters. (杜ピルマブ, 杜ピル马布, 杜ピ尔马布)
French	• The translated CEA abstract did not provide specific language of time horizon and transition probabilities were translated awkwardly • RWE abstract, statistical languages, and endpoints were confusing • Tranlsated CEA abstract didn't cover the main concept of the abstract • Specific language was not translated well • All specific language that was related to the topic had to be redone, poor quality • Unable to grasp the main theme of the abstract based on the translation provided
German	• In general, good enough to understand content of the abstract • A native speaker would have phrased some of the sentences differently • Tiny mistakes show that it is not a translation made by a human • Certain technical terms are not usually translated (ie, QALY)
Italian	• Overall, the translation conveys the general meaning of the abstracts • Specific language was not always properly translated, particularly in the CEA abstract (eg, Cost-effectiveness translated with "utlitá' economica" and "outcomes" translated with "risultati") • Some technical terms were not translated and this reflects current practice in Italian HEOR content (eg, "follow-up" and "panel") • An Italian reader might likely discern that the abstracts were generated by AI due to the improper use of verbs or tenses
Japanese	• Overall, inconsistent and unreliable translation were most common in CEA, due to the limited evidence in Japan. Translations were more accurate in other topics • Some technical terms were inconsistently translated into either into Kanji or Katakana in an article. This can lead to frequent fluctuations in the translation of these terms, resulting in less clarity and readability of the document • An interesting negative point was that terms which should be written in Katakana were instead translated into Kanji • The term "clinically meaningful" was translated differently across articles

CEA - cost-effectiveness analysis; N/A - not available; PRO - patient-reported outcomes; RWE - real-world evidence; SLR - systematic literature review.

DISCUSSION

- This study adds to the ongoing research on the benefits and potential uses of AI tools to support non-native English speakers in comprehending scientific literature originally written in English.
- Previous assessments of ChatGPT's translation ability have highlighted the potential limitations in translating underrepresented languages, leading to translation errors and illogical translations.[7]
- Another study highlighted that AI can support the scalability and speed of translating large documents, but it is still unable to interpret context, cultural nuances, and idiomatic expressions found in non-English sources.[8]
- A key strength of this assessment was the participation of 5 HEOR researchers, covering five different languages and various research disciplines within HEOR.
- One limitation of this study was that the same AI tool was used for the reverse translation step of our process. A potential reason for the high consistency with the original abstract.

CONCLUSIONS

AI showed promising capabilities in translating HEOR publications from English to other languages. This capability could help streamline the dissemination of research findings across global markets. Nevertheless, while AI tools have made notable progress, there remain opportunities for improvement. This includes correct use and translation of HEOR specific terms and better handling of complex economic and medical language, particularly in underrepresented languages. Enhancing these areas could further support HEOR professionals in leveraging AI-driven translations.

REFERENCES

1. Steigerwald E, Ramírez-Castañeda V, Brandt DYC, et al. Overcoming language barriers in academia: machine translation tools and a vision for a multilingual future. BioScience. 2022;72(10):988-998. doi:10.1093/biosci/biac062
2. Longo R, Ghinelli F, Torelli F, et al. OMNI® surgical system versus iStent inject® with concomitant cataract surgery for the treatment of mild-to-moderate primary open-angle glaucoma in the United States: a cost utility analysis. Expert Review of Ophthalmology. 2023/03/04 2023;18(2):135-142. doi:10.1080/17469899.2023.2193685
3. Stanford RH, Blaiss M, Bleecker ER, et al. Real-world effectiveness of dupilumab in biologic naïve patients with asthma: analysis from the US advantage study. Conference Abstract. American Journal of Respiratory and Critical Care Medicine. 2023;207(1)doi:10.1164/ajrccm-conference.2023.C31
4. Cha-Silva AS, Gavaghan MB, Bergroth T, et al. Effectiveness of nirmatrelvir-ritonavir for the prevention of COVID-19—related hospitalization and mortality: A systematic literature review. American Journal of Therapeutics. 2024;31(3):e246-e257. doi:10.1097/mjt.0000000000001744
5. Mercuri E, Darras BT, Chiriboga CA, et al. Nusinersen versus sham control in later-onset spinal muscular atrophy. N Engl J Med. 2018;378(7):625-635. doi:10.1056/NEJMoa1710504
6. Armstrong EP, Malone DC, Franic DM, Pham SV, Gratie D, Amin A. Patient experiences with clostridioides difficile infection and its treatment: A systematic literature review. Infect Dis Ther. 2023;12(7):1775-1795. doi:10.1007/s40121-023-00833-x
7. Deck A. ChatGPT faces challenges in global language testing. Rest of World. Published 2023. Accessed October 18, 2024. https://restofworld.org/2023/chatgpt-problems-global-language-testing/
8. Moneus AM, Sahari Y. Artificial intelligence and human translation: A contrastive study based on legal texts. Heliyon. 2024;10(6):e28106. Published 2024 Mar 14. doi:10.1016/j.heliyon.2024.e28106