National Council on Interpreting in Health Care

Guidance For Healthcare Organizations Evaluating the Potential Use of AI-generated Interpreting

Author: Cynthia E. Roat, MPH; reviewed and ratified by the NCIHC Board of Directors, July 2024

Introduction

The National Council on Interpreting in Health Care is a multidisciplinary non-profit organization whose mission is to promote and enhance language access in health care in the United States.

Purpose of this guidance

As of mid-2024, healthcare organizations in the United States are being approached by companies offering to replace their interpreting and translation services with services provided entirely by artificial intelligence (AI), that is "machine interpretation produced by an AI software during real-time human communication without any input of humans during this interaction." (SafeAI guidance, Section 2.3.)

While the NCIHC recognizes that AI has the potential to improve the quality and efficiency of human-generated language services, we have significant concerns about a transition to exclusively AI-generated services at this time. We recognize that AI technology is rapidly advancing, and that some current limitations may change as technology improves.

Following the example of the non-profit initiative SafeAI and based on their guidance released in late June 2024, this document contains a list of questions that healthcare organizations should ask when considering the use of AI-generated interpreting products. These questions aim to ensure high quality language services that support patient safety, good patient outcomes, compliance with legal and accreditation standards, and financial efficiency.

Questions for healthcare organizations to answer internally:

1. Patient and Provider Choice
   How will we provide a linguistically and culturally unbiased opportunity for patients and providers to accept or decline the use of AI-generated interpreting every time it is used, with the option to choose a qualified human interpreter instead?

2. Data Collection and Storage
   AI-generated interpreting requires that the conversation be recorded, converted to text, translated and voiced in a second language. How will we provide an opportunity for patients and providers to opt into or opt out of data collection and storage without penalty?

3. Workflow Integration
   How will these tasks integrate into staff workflow?

4. Protocol for Switching to Human Interpreters
   What will be the protocol to switch to a human interpreter in the middle of a session if necessary?

5. Documentation
   How will we document the use of an AI-generated interpretation?

6. Sight Translation
   How will we handle sight translation since the AI will not be able to provide that service?

7. Tracking errors
   How will we track interpretation errors to measure the quality and safety of patient-provider communications to improve the technology?

8. Regional dialects
   Are we currently able to identify regional dialect variations in our patient population to identify the right AI interpretation product?

Questions to ask the vendor of the AI services:

Available languages

1. Languages
   Healthcare organizations provide services to speakers of many different languages, including ASL.
   a. What languages are available for bidirectional interpreting?
   b. For each of these, what are the performance evaluation metrics; that is, how accurate to intended meaning is the AI?
2. Low Resource Languages
   a. How are you able to provide services for Low Resource Languages, that is, languages for which there is too small a corpus to adequately train an AI?[1]
3. Regionalism
   a. Is this AI localized to any regional dialects and accents?
   b. How does it manage speech in dialects other than the one(s) it was trained on?
   c. How will the AI identify and mitigate the bias introduced by the specific dialects, variants, and accents used for training the AI?

Compliance with National Code of Ethics and Standards of Practice

Qualified human interpreters comply with the National Code of Ethics and Standards of Practice for Interpreters in Health Care. These questions deal with some of those tenets.

4. Accuracy
   a. Accuracy to meaning
      How well does the AI render messages accurately for meaning, without adding, omitting or substituting?
      i. Has this AI been trained on discourse and terminology pertinent to health care in particular?
      ii. How does the AI's confidence scoring compare to that for a session interpreted by a qualified human interpreter in each language?
      iii. How well does the AI recognize and replicate register, style and tone?
   b. Pre sessions
      i. Qualified human interpreters conduct a pre-session before an interpretation in order to understand the goals for the encounter. How can AI ensure alignment with the goals of the encounter?

[1] Low Resource Languages" can be understood as less studied, resource scarce, less computerized, less privileged, less commonly taught, or low density, among other denominations." Source: Low-resource Languages: A Review of Past Work and Future Challenges https://arxiv.org/pdf/2006.07264

ii. A pre-session also assures that all parties address each other, understand that everything will be interpreted, and pause often to allow accurate consecutive interpreting. How does this AI accomplish this?

c. Managing the flow of communication
Qualified human interpreters primarily use consecutive interpreting, switching to whispered simultaneous interpreting when one speaker does not pause or when speakers of one language are talking together while those of another are excluded. How would the AI handle this?

d. Correcting errors
Qualified human interpreters monitor their own production and correct errors. How does the AI self-monitor?

e. Clarify misunderstandings transparently
Human speech can often be unintelligible or contradictory. Qualified human interpreters will intervene to clarify in order to assure accuracy to intended meaning. When they do so, they make sure both speaker and listener are aware of what they are doing. How does this AI accomplish this?

f. Identification of lack of understanding
Qualified human interpreters monitor listeners for signs of understanding or lack thereof, both through audible signs and body language. How will the AI do this?

g. Who will be held liable for errors in interpretation?
Human interpreters and the companies that employ them are responsible for the quality of the interpretation they provide. If the AI provides an inaccurate interpretation, who is liable?

5. Confidentiality
Qualified human interpreters hold all private healthcare information in strict confidence in compliance with HIPAA. A slip may expose the data of one patient. Does the AI record interpreted sessions? If so, where is the data held and for how long? How is this data protected? What is the data used for? What are the implications of a data breach in which thousands of patients' data is exposed?

6. Impartiality
Qualified human interpreters constantly monitor themselves for personal biases, to assure that these are not reflected in their interpreting. AIs reflect the biases of their training sets. How have you addressed this concern?

7. Culture
Qualified human interpreters identify and call attention to likely misunderstandings based on unaddressed cultural differences. How will the AI accomplish this?

8. Respect
Qualified human interpreters adjust respectful speech in the source language to reflect the same respect in the target language. In addition, qualified human interpreters encourage participants to address each other, even if they start out by saying "Tell him…" or "Ask her…" How will the AI accomplish these tasks?

9. Advocacy
When a patient's health or well-being is at significant immediate risk, human interpreters will speak up on behalf of the patient. How will the AI accomplish this?

10. Continuing Education
    As healthcare progresses, new terminology, procedures and technologies are introduced. How will the AI be kept up-to-date on these new aspects of healthcare? Who will be responsible for that process?

Compliance with Federal regulations

11. How does this AI comply with Section 1557 of the PPACA rule (§ 92.201 Meaningful access for individuals with limited English proficiency) that requires review by a human linguist when the material to be translated is critical to the rights, benefits, or meaningful access of an individual with limited English proficiency, when accuracy is essential, or when the source documents or materials contain complex, non-literal or technical language? How will there be human oversight for an on-demand service?

Quality assurance

12. Incorporation of a human interpreter
    If a provider initiates a conversation through this AI and then must switch to a qualified human interpreter due to inadequate or inappropriate AI responses, how will this be tracked and communicated?

13. User Feedback
    Can end users "grade" the AI interpreter? If so, how do these end-user "satisfaction scores" compare to those received by qualified human interpreters?

14. Bias Audits
    If this AI product been audited for racial and other biases, what were the results?

15. Incorporation of Feedback
    How is feedback incorporated into the machine learning of this AI? At what points can the tool be adjusted or taught to adjust? Do you offer clients a mechanism to provide direct feedback to this system?

16. Grievance Protocols
    If we are not satisfied with the quality of the AI interpretation, what are the protocols for submitting a grievance?

Ownership of information garnered by AI

17. Who will own the information garnered by the AI in the course of interpreting? If it is the vendor, what will be done with that information?

Additional concerns

As you consider the usefulness of an AI-generated interpreting system, consider as well that the language proficiency of AI platforms is constrained by the data they have been trained on and may not be as nuanced as a human interpreter, especially regarding idiomatic expressions and the diverse language regionalisms, culturally and contextually dependent, of each language. In addition, AI platforms inherently lack the intrinsic understanding of ethical principles that a qualified human interpreter possesses. Finally, qualified human interpreters often play a crucial, but understudied role not only in ensuring effective communication but also in building empathy and trust in healthcare settings, aspects of care that AI will not be able to fully replicate.