CSA Research

Summary of "Perceptions on Automated Interpreting"

Results of a Large-Scale Study of End-Users, Requestors, and Providers of Interpreting Services and Technology

Study Conducted on Behalf of the Interpreting SAFE-AI Task Force

March 2024

Hélène Pielmeier

Contents

List of Figures

Introduction

The Interpreting SAFE-AI Task Force commissioned independent market research company CSA Research to develop, run, and analyze a large-scale perception study of end-users, requestors, and providers of interpreting services and technology. The goal of the study was to capture current perceptions about spoken and signed artificial intelligence for interpreting, with a focus on the US market.

This summary presents a high-level view of core findings. Refer to "Perceptions on Automated Interpreting" for the full analysis.

> ***Note:*** *Each section in this summary provides details on how to locate more information in the 350-page study.*

Scope and Limitations

Ultimately, the Interpreting SAFE-AI Task Force seeks to establish guidelines for the responsible adoption of AI. Given the absence of previous research and the enormous breadth of the topic, we acknowledge four limitations in the scope of this research:

- **Interpreting-centric.** This preliminary study focuses on language interpretation. The Task Force is interested in later capturing similar perspectives regarding transcription, captioning, and subtitling as these automated solutions rely on the same technology backbone.

- **Spoken language.** The original intent was to also capture perceptions about both signed and spoken languages. However, the end-user portion of the study ended up focusing on interpreting in general. An independent "Advisory Group on AI and Sign Language Interpreting" was formed to examine aspects specific to automated interpreting into sign languages and its effects in the Deaf community. This group published its results at https://safeaitf.org/deafsafeai/ as a supplement to this report.

- **US-focused.** The Task Force chose to primarily target respondents in the United States due to the group realizing it couldn't handle the scope of a larger-scale study. While we graph responses for in the US versus outside the US, we didn't capture the

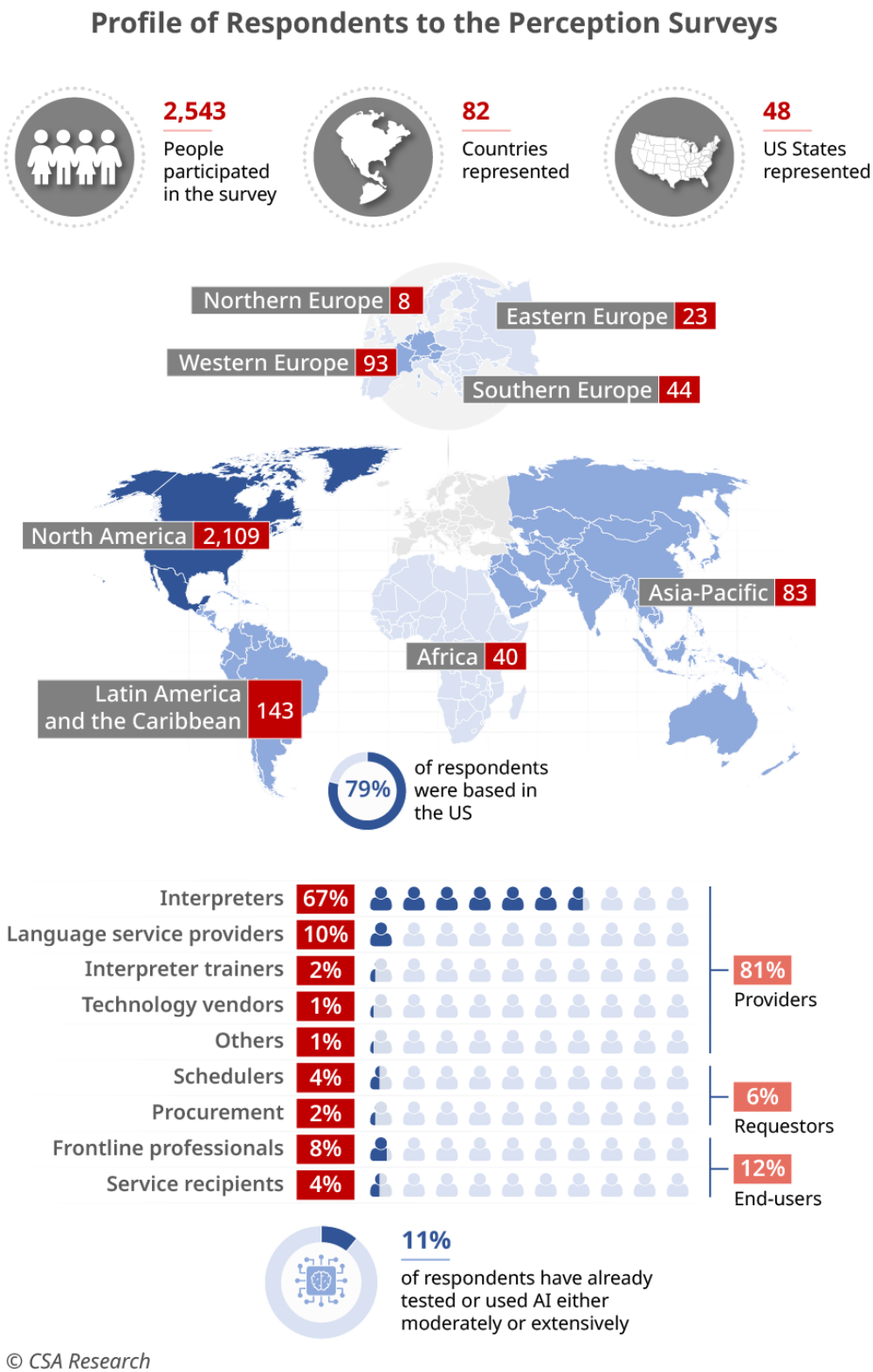
same depth of data for the latter geographies. Separate research will be necessary to capture viewpoints for other countries or regions.

- **More work to be done.** This report focuses on capturing current perceptions regarding the use of AI in interpreting. On their own, the findings of this report are not sufficient to develop guidelines. Further research will be required by use case scenario and industry to establish a strong framework – research that will involve in-depth interviews with consumers and technology vendors and hands-on experience with the systems. However, documenting the perceptions of the various constituencies is an important first stage of data collection that will enrich follow-on studies.

Respondent Profile

The 2,543 respondents came from 82 countries but 79% were from the US. Two-thirds of respondents (67%) were interpreters, and more than three-quarters of interpreters (77%) worked in health care. This means that the results of this study are highly representative of the perceptions of interpreters – particularly those working in the medical field. This is a high-risk sector where mistakes are not as forgivable as in other domains. As a result, language professionals had an intense negative reaction to the thought of using AI (Respondent Profile in Chapter 2).

Figure 1: Profile of Respondents to the Perception Surveys



This huge number of responses from interpreters working in high-risk areas means that the view of this study does not represent the market as a whole. The number of end-users was too low to capture enough nuances on when service recipients and frontline professionals find some benefits – especially as 43% of frontline professionals also worked in the medical field themselves. The low response count for end-users ties to the grassroots nature of the recruitment effort (The Respondent Profile Led to an Emphasis on Negative Points in Chapter 15).

> ***Note**: We also had difficulty engaging with companies that build AI interpreting systems. We believe it was tied to a fear of what might result from the study in terms of usage recommendations. The majority of technology vendors who participated were actually not interpreting AI vendors; they more closely fit the profile of interpreting companies with their own (human) interpreting delivery platform. Most focus on AI captioning, subtitling, or transcription (Profile of Providers in Chapter 2).*

Thank You

CSA Research wants to thank the people and organizations who donated time or resources to scope, fund, test, and translate the survey. Special thanks go to financial donors: Akorbi, AMN Healthcare, Boostlingo, Certified Languages International, Cesco, Cross-Cultural Communications, Lango, LanguageLine, MasterWord Services, National Council on Interpreting in Health Care (NCIHC), PGLS, Sorenson, The Language Group, Translation Station, Universal Language Services, and WP Rivers Associates.

Core Findings

The rapid evolution of AI is reshaping the language industry, prompting questions about its impact on the interpreting sector, language and communication access, and the future of the interpreter profession.

Potential benefits include increased access to language services, reduced wait times – thanks to round-the-clock availability and no need to schedule interpreters – and lower costs. However, in the absence of government guidelines, organizations using interpreting services and technologies can make their own implementation decisions – good or bad. Respondents across the range of roles fear that uninformed users across commercial, government, and non-profit organizations will make the wrong call because they view the technology as a cost-cutting opportunity and assume it delivers quality. And as a result, end-users would pay the price through health issues, negative financial impact, or loss of freedom.

The Interpreting SAFE-AI Task Force's goal is to build currently missing best practices and safety guidelines to ensure that AI is used ethically and responsibly for language and communication access. The group strives to help the market evolve from tech buzz to ensuring meaningful access for individuals with limited proficiency in English – or any other language.

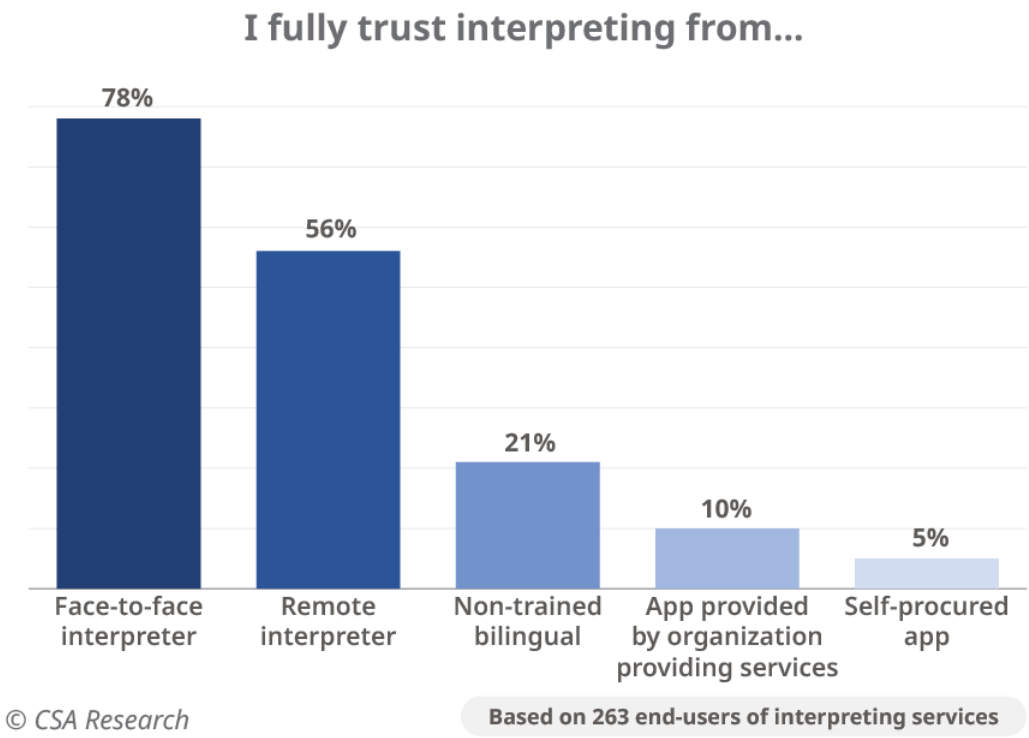
This summary synthesizes findings about: 1) trust; 2) impact of experience with AI; 3) accuracy limitations; 4) AI appeal; 5) challenges with AI solutions; 6) disconnect on core benefits; 7) requirements for deployment success; 8) the impact of ethics; 9) the disconnect in applicability scenarios; and 10) the effects on the role of the interpreter.

Trust in AI is Low

Only 10% of end-users of interpreter services "fully trust" automated interpreting from apps provided by the organization rendering services. And that percentage drops to 5% when the output comes from a free or low-cost solution found in an app store. While these numbers are very low, they are not significantly worse than those for non-professional interpreting done by friends, family members or coworkers (fully trusted by 21% of end-users). Remote telephone or video interpreting (fully trusted by 56%)

also lags behind in-person interpreting (78%) by 22% (Trust by Source of Interpreting in Chapter 10).

Figure 2: Full Trust in Interpretation from Different Sources



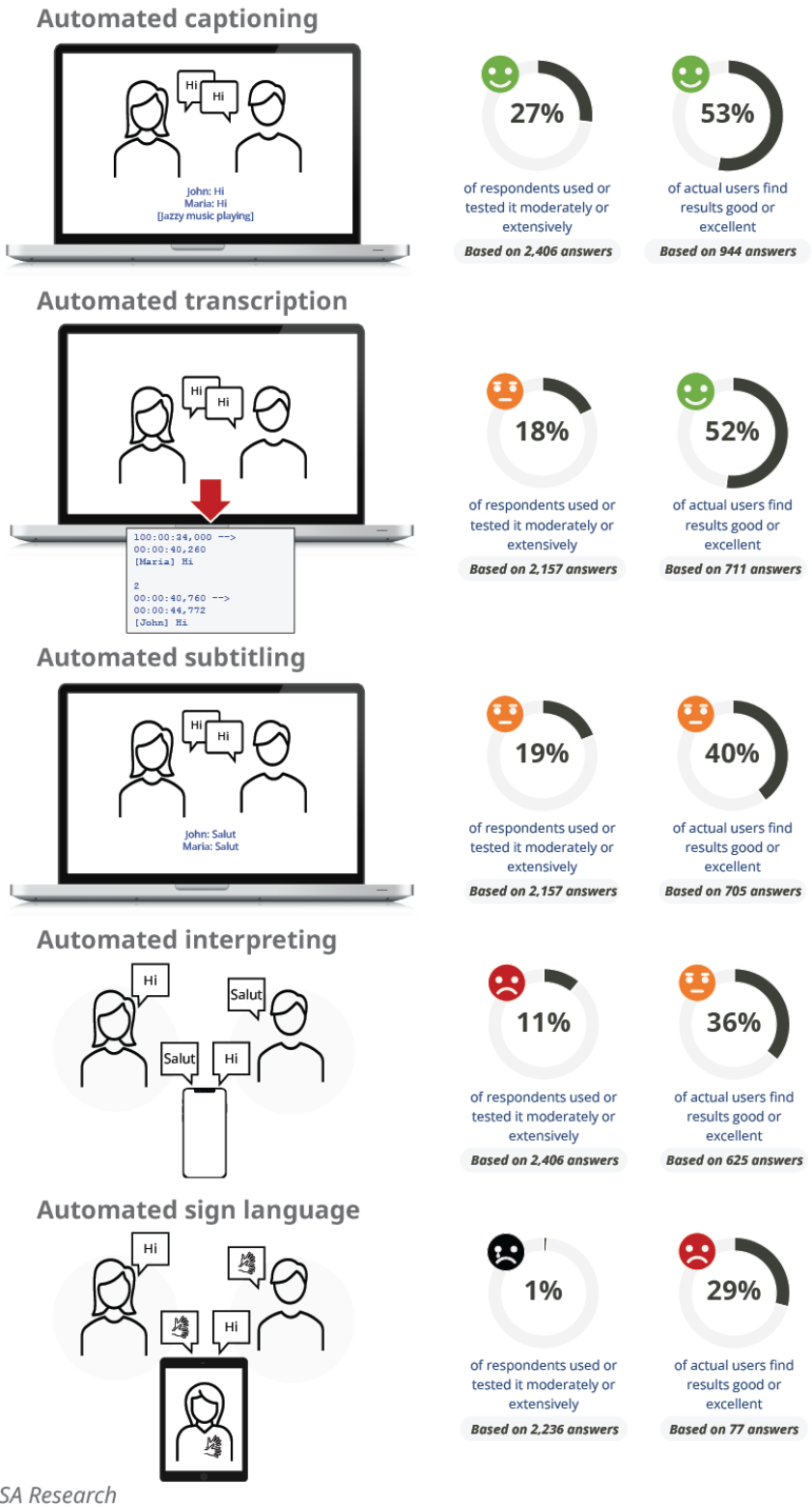
So, what level does AI have to reach to be considered usable? Although there is no clear answer, what respondents fear is the loss of vital information that could affect a ruling, a diagnosis, or next-step actions – each of which can have huge financial and human costs. To trust AI, respondents must feel like potential consequences will not be more severe than in the current situation of working with a human interpreter or not providing interpreting at all (Negative Outcomes Feared by Respondents in Chapter 9).

Part of the struggle with achieving greater trust also comes from concerns about how technology vendors store data, keep it confidential, and prevent bias. Respondents worry about potential data breaches and misuse of technology. They want regulations and safeguards (Confidentiality and Bias in Chapter 9).

Those Without AI Experience Underestimate AI’s Capabilities

Only 11% of participants had extensive to moderate experience with automated solutions for spoken interpreting and a mere 1% claimed the same for signed language solutions (Experience Using Automated Solutions in Chapter 5).

Figure 3: Five Automated Services: Experience and Perceptions on Quality

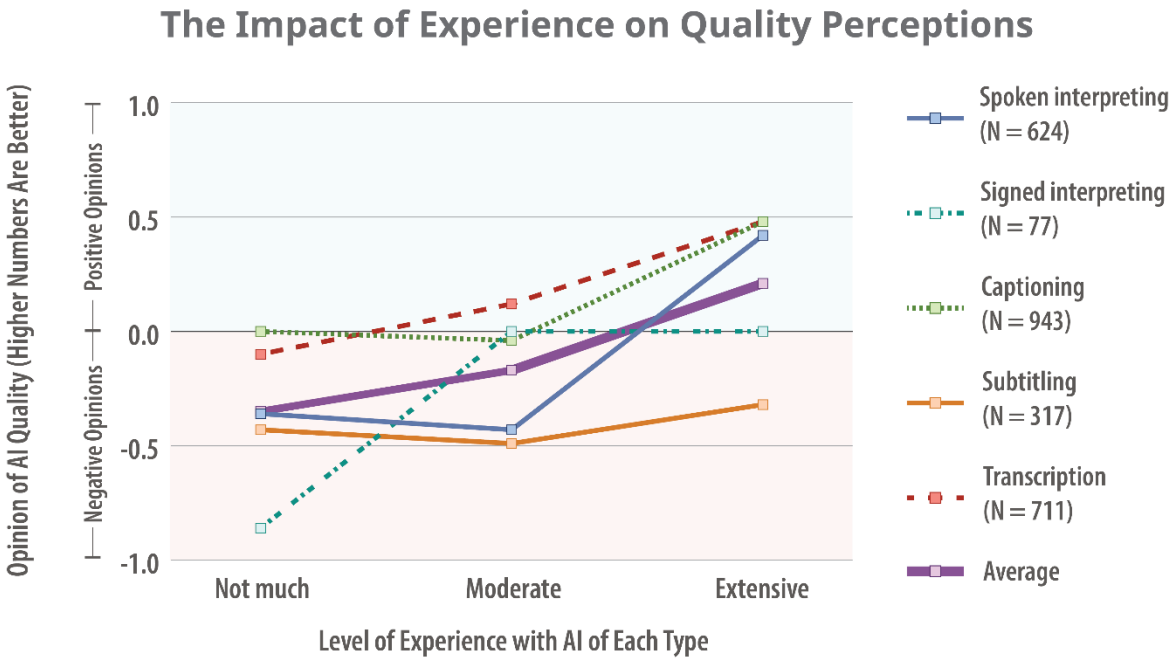


© CSA Research

This means that the vast majority of survey takers judged automated interpreting's capabilities without experiencing for themselves the caliber of output from the latest technologies. If people are really opposed to AI but have never experimented with a solution, their feedback is influenced by their belief system and is not based on the actual capabilities of the technology (The Respondent Profile Led to an Emphasis on Negative Points in Chapter 15).

The more experience people had, the more positively they thought about AI's capabilities – or the more they have an interest in AI, the more likely they were to try it. The effect is relatively modest, but it is real: Internet or exposure to tools helps improve perceptions (Thoughts on Quality of Automated Solutions in Chapter 6).

Figure 4: The Impact of Experience on Quality Perceptions



© CSA Research

Accuracy Is the Main Element in Gauging Capability

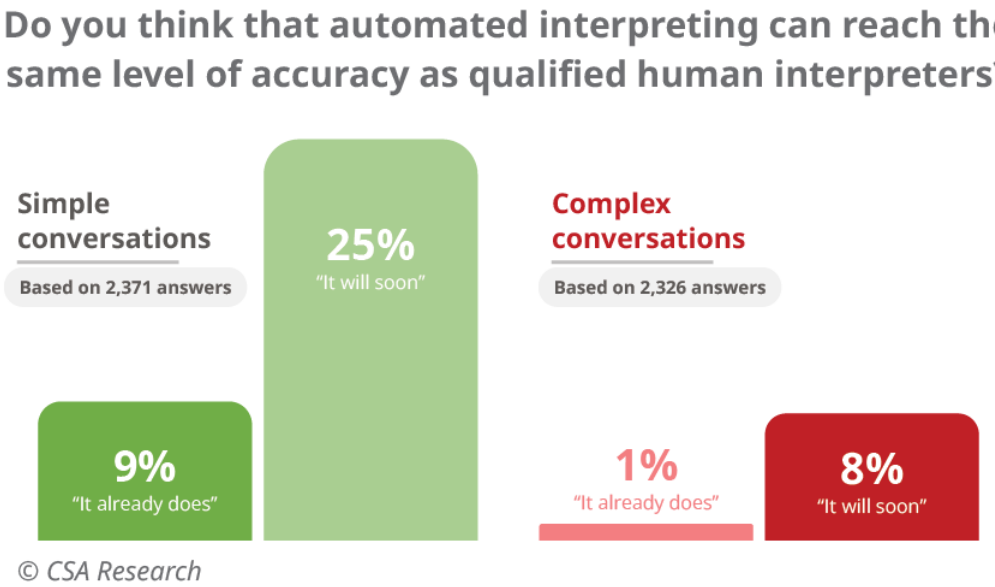
While AI is getting a lot of press, it’s not ready to “take over” the whole market.

On the signed language interpreting front, text-to-sign and sign-to-text technology remains very basic, and is not suitable yet for bidirectional conversations. Every language – and sometimes even countries sharing the same spoken tongue – has a corresponding unique signed language – yet few are covered by existing AI. Signed language technology is currently only available for prepared or prerecorded content rather than for on-the-fly use. In addition, much of the nuance that matters in signed languages cannot yet be modeled in AI systems (Solutions Related to Signed Language in Chapter 4).

Spoken language interpreting is further ahead thanks to great advances in voice recognition, machine translation, and voice synthesis. However, the technology compounds errors from each of these process steps so results are good overall but remain imperfect and unpredictable. In some contexts, you can gloss over imperfections, while in others, they can lead to significant negative outcomes that carry too much risk (Solutions Related to Spoken Language in Chapter 4).

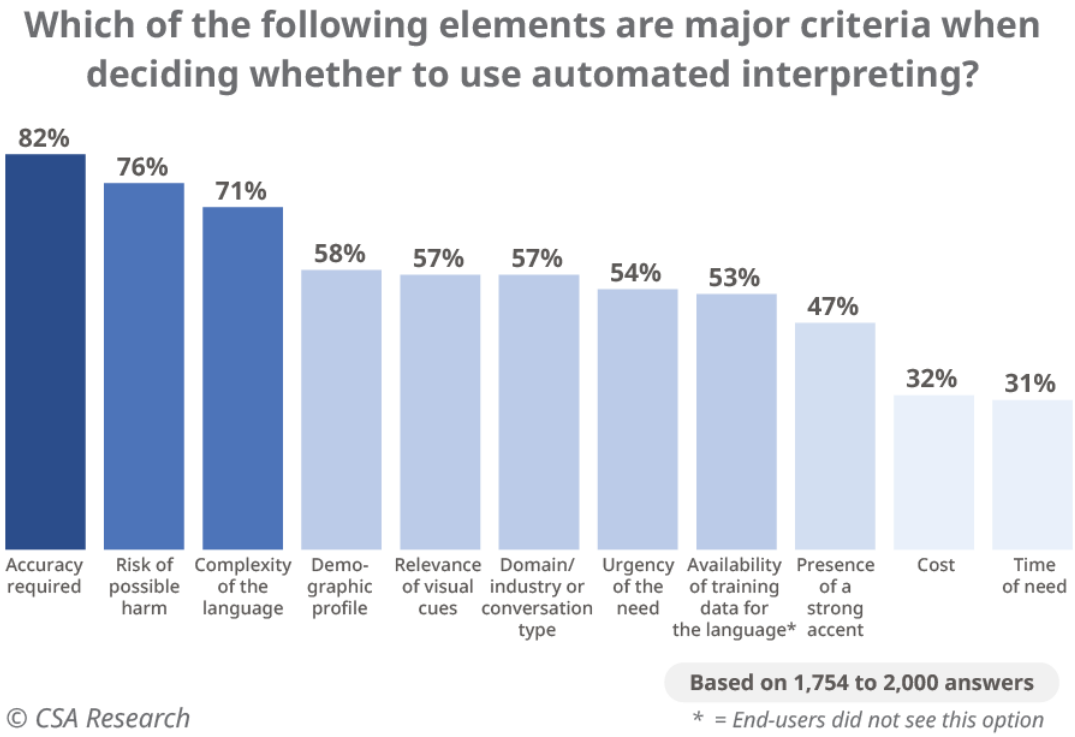
We asked respondents to contrast opinions on the amount of time it would take for automated interpreting to achieve the same level of accuracy as qualified human interpreters. About one-third of respondents (34%) stated that human parity already exists (9%) or will soon do so (25%) for simple conversations. That number naturally dropped to 9% for complex conversations. The challenge comes in defining what constitutes simple versus complex conversations, especially as few conversations stay at the “simple” level throughout (Perceptions of Accuracy for Automated Interpreting in Chapter 6).

Figure 5: Timeline to Reach Human Parity



When we examine criteria to adopt when determining whether to use automated interpreting, the top criterion across respondents was the degree of accuracy required from the session. Fully 82% of respondents strongly believed accuracy was a major decision element. Responses were quite consistent across audiences except in one important case – service recipients were less concerned about accuracy than other roles (Accuracy Required in Chapter 12).

Figure 6: Major Criteria When Deciding to Use Automated Interpreting



Overall, respondents want proof of accuracy to gain trust in AI solutions before they can comfortably consider it for deployments. They call on technology vendors to supply these proofs and on third parties to validate their claims (Accuracy Requirements in Chapter 14).

Even though no technology vendors have claimed perfection in their systems, total accuracy is what many of our respondents expect from AI interpreting. As a result, current outputs fall short of expectations. However, these expectations may not be in line with the goals of technology vendors who primarily focus on low-impact language and communication access where the alternative to automated interpreting is no interpreting at all – cases when it is not economically or practically viable to supply human interpreting (The Technology Is Not Perfect in Chapter 15).

Respondents tended to consider the question of automated solutions as a black-and-white situation – either all human or all automated interpreting, failing to see the shades of grey in between. And since the technology doesn't meet their idea of quality,

many want nothing to do with AI. However, those who could see shades of grey tended to lean toward the opinion that in high-risk situations, automated interpreting should not even be an option and that AI should instead be used to help human interpreters working in these complex scenarios. But, in some low-risk encounters, automation could play a role (The Technology Is Not Perfect in Chapter 15).

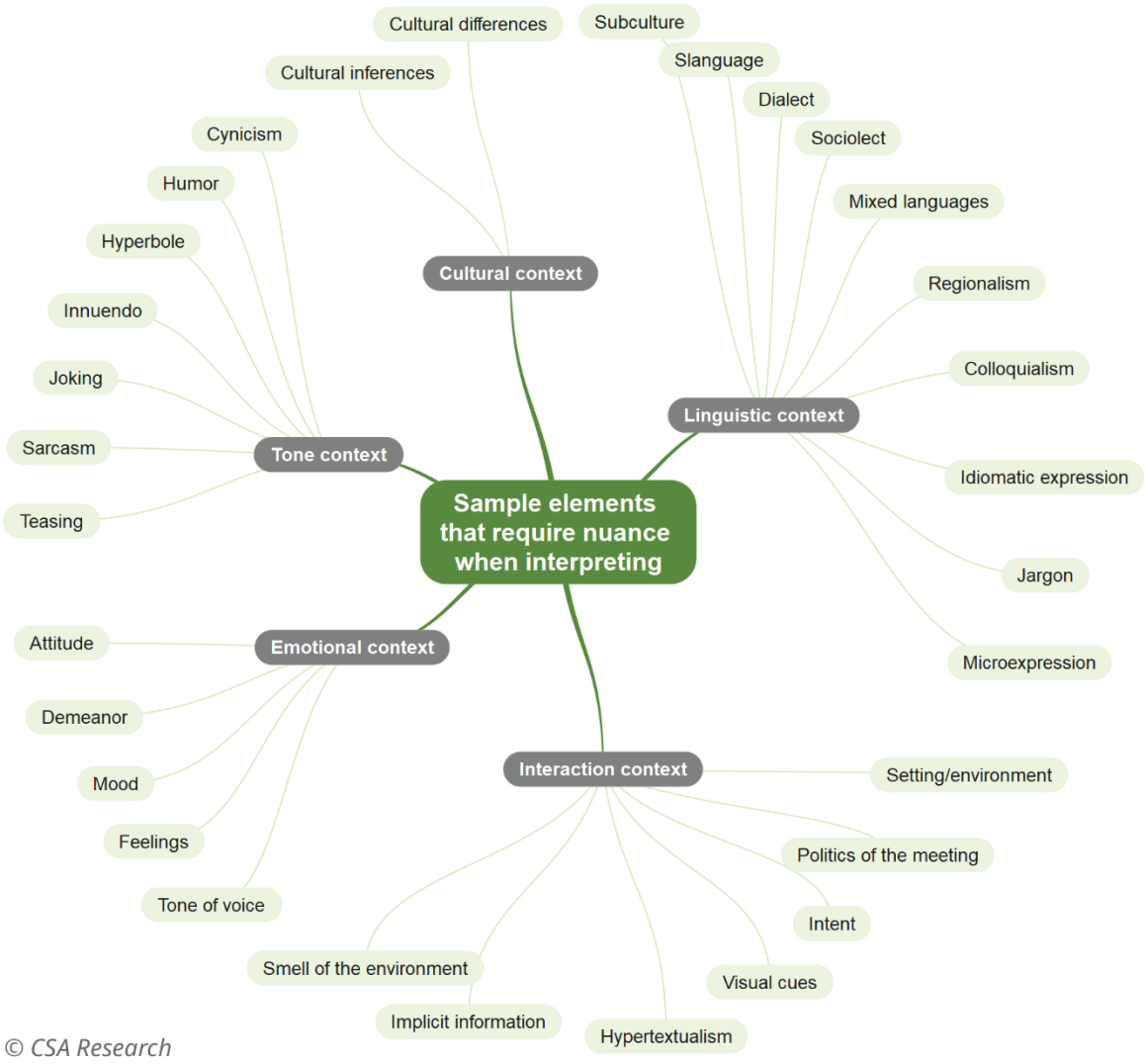
Interpreters Want Synergy, Not Replacement

Language professionals are significantly concerned about the impact of automation on human interpretation. Many of them feel threatened by AI, believing it could lead to job loss, reduced income, and the devaluation of their skills and expertise (The Existential Angst in Chapter 3).

Many believe that AI cannot replace human interpreters due to its limitations in understanding meaning, emotion, cultural differences, and body language (Comments from Respondents in Chapter 9).

Respondents emphasized AI’s inability to deal with subtleties and ambiguities. Interpreting requires the ability to deal with the context for language, culture, tone, emotions, and interaction background. AI can’t capture visual cues, cultural inferences, understand the participant mood, or handle regionalisms. In short, respondents fear that because AI can’t see nuance or read between the lines, service recipients have too much to lose when the machine doesn’t consider context and culture (Obstacles Resulting from Context and Culture in Chapter 9).

Figure 7: Sample Elements That Require Nuance When Interpreting



Similarly, the language people speak is not textbook perfect. Neither are situations in which interpreting occurs. Interpreting for someone who uses poor grammar, has a speech impediment, or is in an active psychotic episode – to name just a few potential challenges – requires some serious skills. The reason that humans remain the better choice is that they can provide language accommodations for unusual or unexpected situations by adapting what and how they interpret using their knowledge and common sense (Challenges Dealing with Imperfect Scenarios in Chapter 9).

Figure 8: Sample Imperfect Scenarios That Require Adaptation When Interpreting

Respondents also fear the deterioration of language services, resulting in discrimination for service recipients. If language or communication access decreases or is reduced in quality, it is likely to contribute to or exacerbate existing problems with discrimination (The Impact on Interpreting End-Users in Chapter 3).

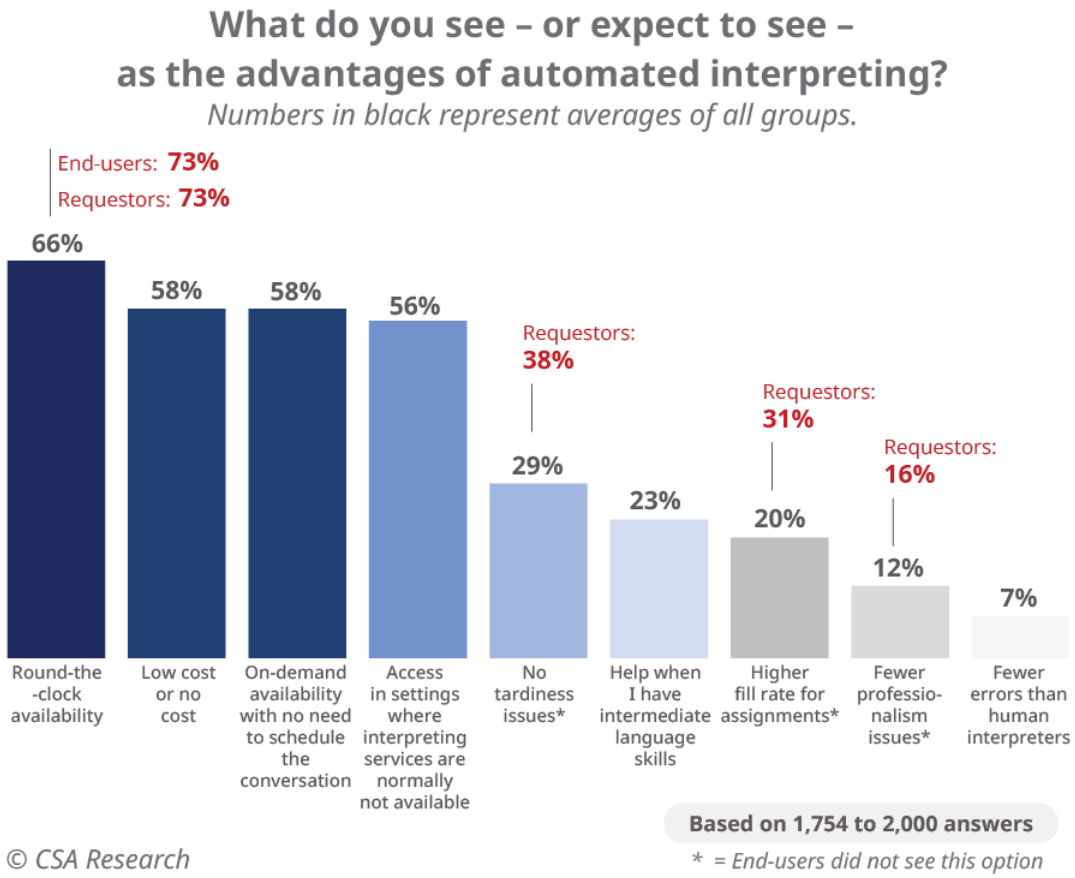
As a result, many professionals believe that AI should be used as a complement to – rather than a replacement for - human interpreters, particularly in high-stakes situations. This is where computer-aided interpreting technology can augment the ability of language professionals by giving them "superpowers" – such as memory

assistance and terminology available in real-time, but only if developers can resolve thorny challenges with cognitive overload (AI as an Aid to Interpreters in Chapter 3).

The Appeal Comes from Ease of Access, Low-Cost, and No Human Complications

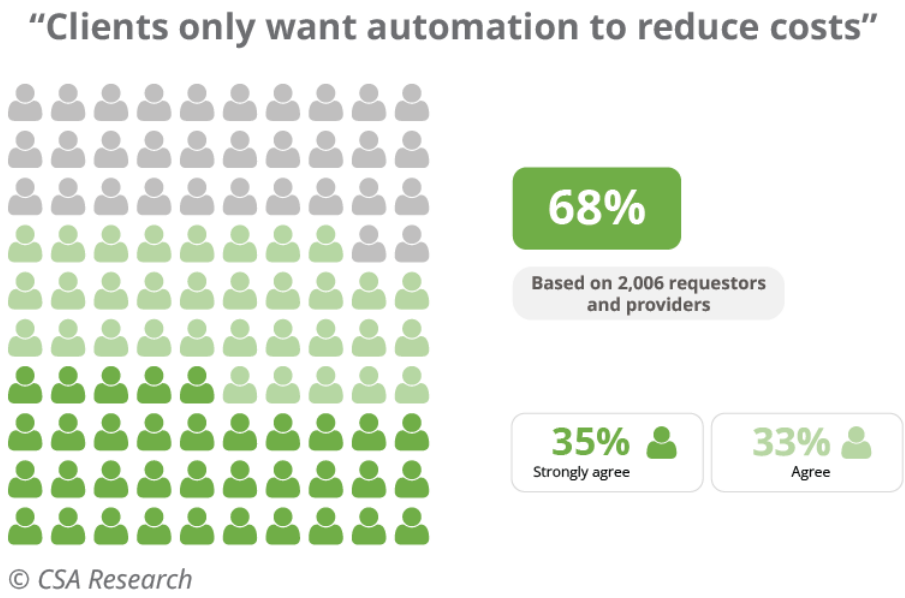
Respondents’ answers saw 24/7/365 on-demand access without phone trees or scheduling as an advantage of AI, with 66% of participants overall who selected this option, and end-users and requestors showing even higher percentages (73%). The tie for second position (58%) was shared between no need to schedule an interpreter and low-cost. However, these three wishes for benefits are not guaranteed in real deployments because AI is available in limited languages, hunting for a physical cart with a video device can be just as frustrating as scheduling, and the costs of mistakes can offset savings (24/7/365 Tops the List of Benefits in Chapter 8).

Figure 9: Advantages of Automated Interpreting



More than two-thirds of respondents (68%) agreed or strongly agreed that requestors only wanted automation as a way to reduce costs. The only respondents who really disagreed with that statement were people in procurement roles who generally view these issues more holistically based on all the elements that go into such decisions. A language professional may not see the cost of the administrative burden involved in scheduling interpreters or what happens when there is a "no show" from either the service recipient or the interpreter (Cost Reduction's Role in Decisions in Chapter 7).

Figure 10: The Cost Factor

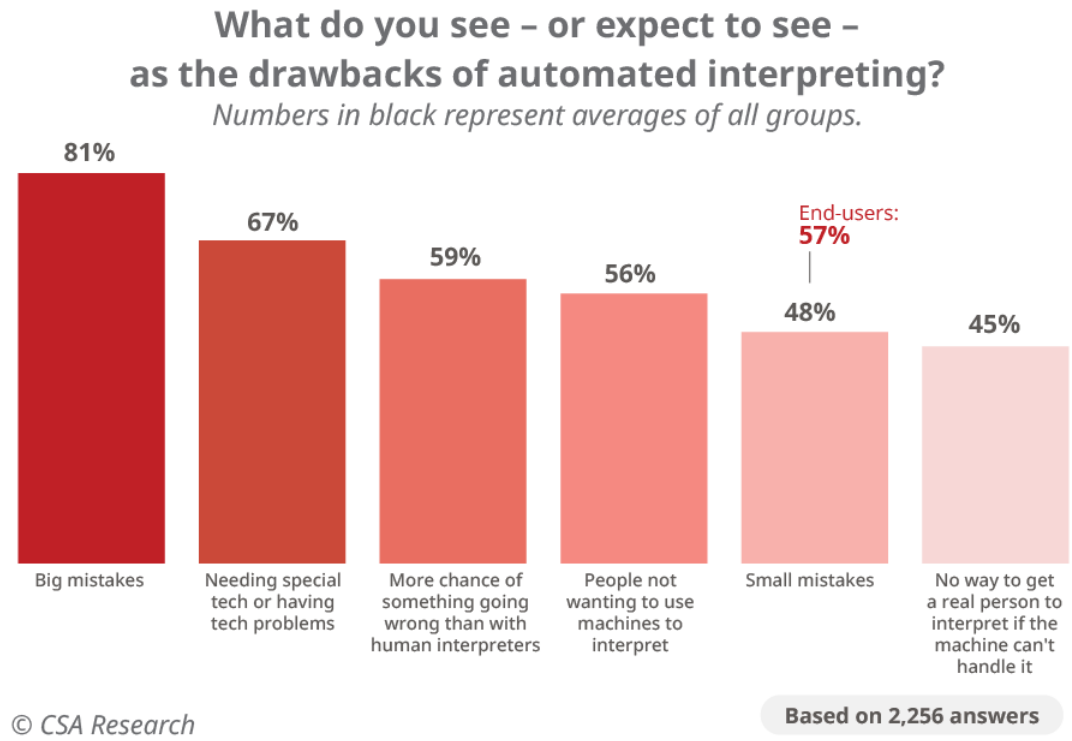


For some respondents, AI is a refreshing alternative when they are not satisfied with the existing options. Requestors had stronger-than-average numbers who saw lack of tardiness as a plus (38% of requestors vs. 29% for all respondents), higher fill rates (31% vs. 20% overall), and fewer professionalism issues (16% vs. 12% overall). End-users exhibited a stronger-than-others reaction regarding errors, with 14% believing AI would lead to fewer errors (vs. 7% overall) (24/7/365 Tops the List of Benefits in Chapter 8).

The List of Challenges Remains Long

The top concern with AI solutions is mistakes. 81% of respondents worry about big mistakes where the main idea might be wrong, or a mistake could cause harm or a legal problem. Nearly one-half of respondents (48%) worry about the smaller ones, when details are wrong – and that number jumps to 57% for end-users (The Top Worry Is Big Mistakes in Chapter 9).

Figure 11: Drawbacks of Automated Interpreting



Respondents underscored the importance of human interpreters, particularly for maintaining human connection in high-stakes situations involving vulnerable populations. They worry about detrimental effects on the quality of language access, potential biases, and discrimination, especially in critical areas like healthcare and legal settings. Some even see AI as a violation of civil rights (Language Access Would Suffer in Chapter 3).

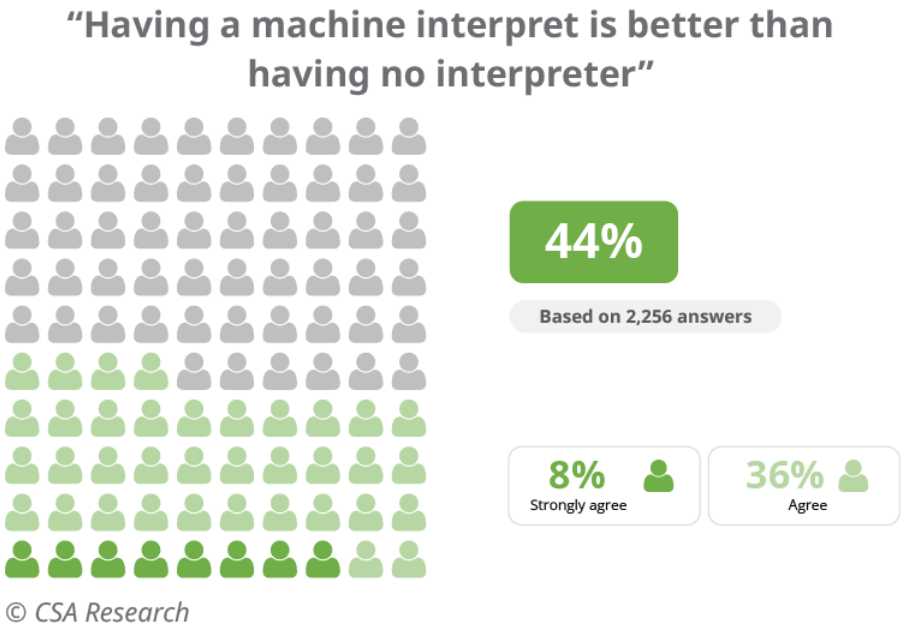
However, the list of concerns doesn’t stop there. Key points include skepticism about AI's ability to fully understand meaning or ask for clarification, technical complications for users, and the lack of human connection. Respondents raised concerns over AI's potential to diminish language richness. Last but not least, respondents fear unforeseen legal and ethical problems (Comments from Respondents in Chapter 9).

Scales Balance Differently Depending on Viewpoint

Another big takeaway from the research is that both ends of the spectrum on a question can hold valid answers.

- **Language and communication access.** Automated interpreting has the potential to simultaneously increase and decrease language and communication access. It can enable the provision of interpreting services when none were available before, but the drop in quality of the interpreting can decrease meaningful language and communication access (Language Access Would Suffer in Chapter 3 and Greater Accessibility of Language Services in Chapter 8).

Figure 12: AI As a Means of Language and Communication Access



- **Languages of limited diffusion and Indigenous languages**. Much of the hope for greater language access revolves around these languages, yet AI models tend to be less developed for them. The dominance of English and European languages in generative AI training data leaves almost 1.7 billion people with no representation, and more than 2.3 billion with minimal representation. Just 10 languages comprise more than 85% of available training data. This situation creates openings for discrimination and biases, as well as highlighting disparities in digital access and

literacy. That means this use case may be wishful thinking and may lead to more issues than in well-trained languages. The determining factor for the availability of automated services for languages of limited diffusion is the same as for human interpretation – it's a question of investment, either in humans or training the machine. (Availability of Training Data in Chapter 12).

- **24/7/365 coverage.** The same applies to the hope of providing round-the-clock availability. Using AI does not guarantee 100% fill rate. For example, telephone and video remote interpreting providers regularly provide services in 150 to 200 languages. If machine interpreting vendors choose to train only the most financially viable languages, automated interpretation's fulfillment rate will be inferior to current remote modalities (Usability When There Is No Interpreter at All in Chapter 7).

- **Privacy.** AI comes with its fair share of concerns over data privacy and potential data breaches. However, having a conversation away from the eyes and ears of a fellow human may have some advantages, especially in small communities where service recipients may personally know the interpreter. Using a machine can be more comfortable when dealing with matters of a highly personal nature, sensitive topics, or anything that requires increased privacy (AI As a Means to Achieve Privacy in Chapter 7 and Confidentiality and Bias in Chapter 9).

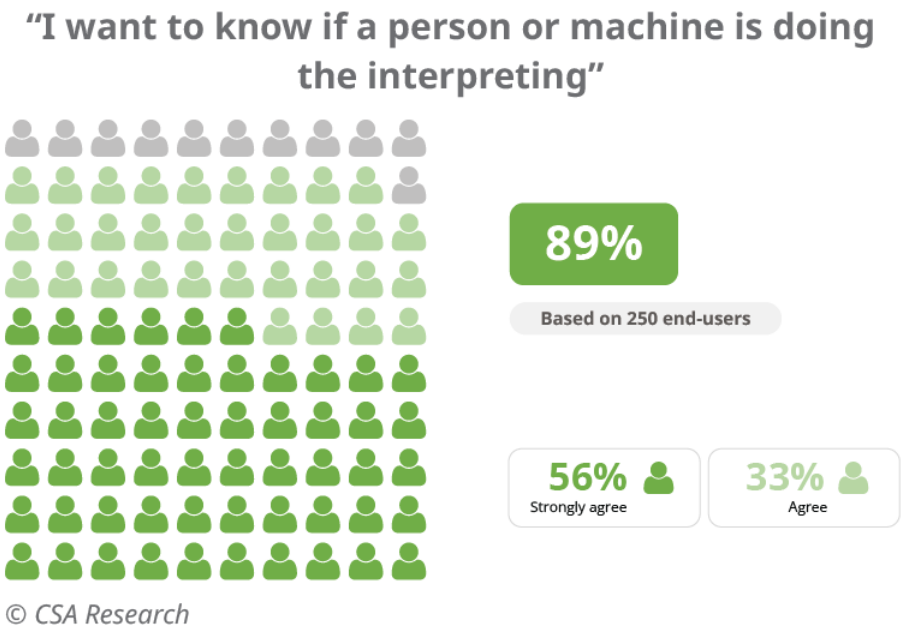
Deployment Success Requires Strong Practices

Ultimately, what is currently missing are best practices for safe deployments that protect users of interpreting services. That covers a variety of different elements.

- **Escalation mechanisms.** Slightly more than one-half of respondents (53%) maintained that the ability to escalate to a human would increase the usefulness of automated interpreting by a lot (26%) or at least a little (27%). For example, apps would include a button to request to talk to an interpreter – telephone, video, or even in-person – if either party felt there was a communication challenge. It's similar to the "press 0" to ask to talk to a customer service representative in phone-based systems (The Impact of Access to a Human for Escalation in Chapter 6).

- **Disclosure.** End-users also don't want to be fooled on whether they deal with a bot or a human. 89% want to be told who's doing the interpreting (Disclosure of Who or What Interprets in Chapter 7).

Figure 13: Disclosure on Who or What Interprets



- **Good audio and relevant scenario.** There are lots of variations in deployment scenarios. The quality of the audio input, whether people enunciate clearly and speak in full sentences, and potential cognitive differences of participants can drastically affect AI's ability to do a good job. Even when two sessions appear to have the same profile on paper, actual conditions can vary greatly (Elements That Affect Output Quality in Chapter 6).

- **Clear accountability and protections.** AI's benefits increase when it includes accountability and data protection protocols, which are essential for the safety of patients and consumers. Some users also call for legal protections and penalties for misuse of AI technology (Lack of Credentials and Accountability in Chapter 9).

- **Guidelines on use.** Respondents call for clear guidelines on ethical use, quality control procedures, and involvement of trained human professionals in the development and implementation of AI interpreting solutions, and a recognition of

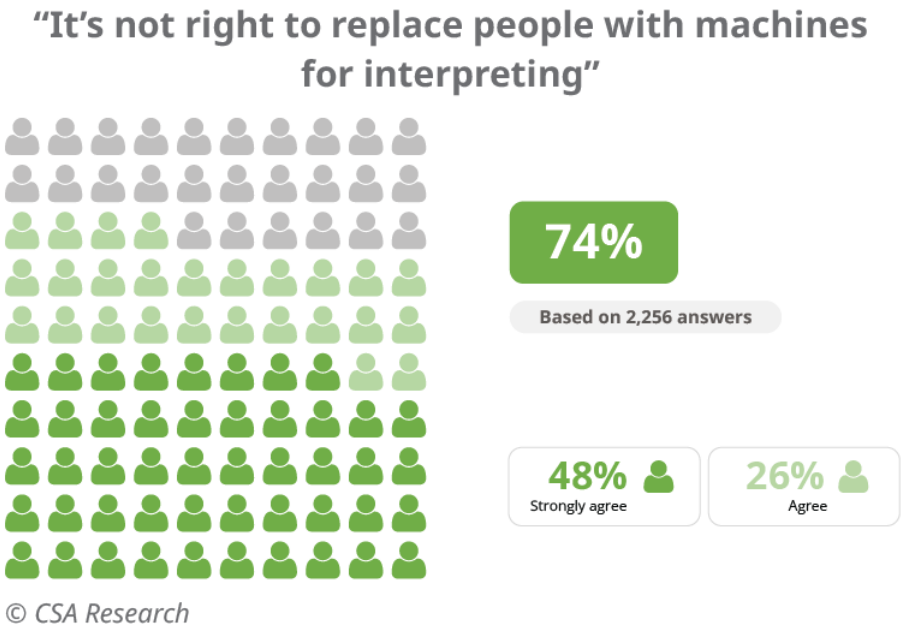
the limitations and potential negative effects. Some respondents fear that things are moving too quickly, and that lack of data makes it premature to establish guidelines. On the other hand, since some organizations are already working with AI, it is urgent to put safeguards in place so that vulnerable populations are not negatively affected by decisions (Prioritize Guidelines Definition in Chapter 15).

- **Exhaustive testing and transparency on testing data.** Participants expressed a desire for AI interpreting technology to be thoroughly tested and proven accurate before widespread adoption in life-critical environments. They call for transparency from AI vendors regarding training materials and data sources (What It Would Take to Increase Trust in AI In Chapter 14).

The Biggest Deployment Hurdle is Ethics, Not Capabilities

At the end of the day, AI is not perfect, but it is usable in a variety of low-stakes scenarios. Respondents overall were more concerned about their perception that it’s not right to use AI than whether AI is capable of doing the job. Many respondents claim that machines “can’t” deliver when they haven’t tested the systems themselves. Their answers are often more about saying that machines “shouldn’t” be tasked with handling interpreting. That is why nearly three in four respondents (74%) either agreed or strongly agreed that it’s not right to replace people with machines for interpreting (The Ethics of Replacing People with Machines in Chapter 7).

Figure 14: The Ethics of Who Should Interpret



Technology vendors were the only ones with fewer qualms about using AI where it can deliver on requirements. The most outspoken anti-AI perspectives came from association representatives and interpreters, but service recipients weren't far behind. The latter were likely experiencing some guilt from accepting interpretation from a machine that took work away from humans (The Ethics of Replacing People with Machines in Chapter 7).

The second big ethical conundrum has to do with who is accountable when mistakes happen. Who will that be when it's an algorithm in the cloud performing the task? Who is at risk of a lawsuit? Can medical professionals risk losing their license to practice? These questions are important to address because human interpreting remains a superior service. If AI were to outperform humans, this ethical issue would drastically change (Lack of Credentials and Accountability in Chapter 9).

Applicability Scenarios Are Full of Disconnects

Many respondents lack technical knowledge of "how the automated interpreting magic happens." This results in significant disconnects between the technology's actual capabilities and how respondents believe AI could be used.

- **Emergencies**. Respondents wish to rely on the expediency of securing AI interpreting. However, because AI interpreting likely requires more back and forth to clarify points, it can make you lose precious time. In emergencies, it may be better suited to simply notify someone that a human interpreter is on their way (Urgent Conversations in Chapter 10 and Emergency Services Use Cases in Chapter 11).

- **Stability of the output.** The challenge with guidelines is also that they assume all technology platforms achieve the same level of accuracy, security, privacy, bandwidth, and usability – which is not the case. Beyond just guidelines, what is necessary sooner rather than later are performance standards for technology vendors, so those who abide by the best practices are not put in the same bucket as those who don't (The Challenge Is to Define Suitable Scenarios in Chapter 15).

- **Delineation of scenarios**. Once we get past respondents who, out of principle, have no interest in seeing AI deployed, we begin to see a consensus that AI can be suitable for low-risk, simple conversations. The big issues are outlining the defining characteristics of such interactions and tracking these characteristics ahead of the session to decide which interpreting method to apply. A use case definition should not just be a domain or a conversation type. Many factors come into play and if any one of them triggers a flag, then recourse to humans will always be preferred. (Usage Scenarios – The End-User Perspective in Chapter 10, Usage Scenarios – The Requestor and Provider Perspectives in Chapter 11, and Use Case Appendix with details on 58 use cases).

  In terms of implementation, the challenge is also that even the simplest interaction at the onset can quickly and unpredictably escalate to a more complex one: Mechanisms need to be in place to bring a human professional into the equation quickly. End-users must be able to trigger an escalation process without negative repercussions to their case. And interpreters must arrive ready for the session with context information and terminology at the ready (The Challenge Is to Define Suitable Scenarios in Chapter 15).

Figure 15: End-Users' Perspectives on the Suitability of Automated Interpreting

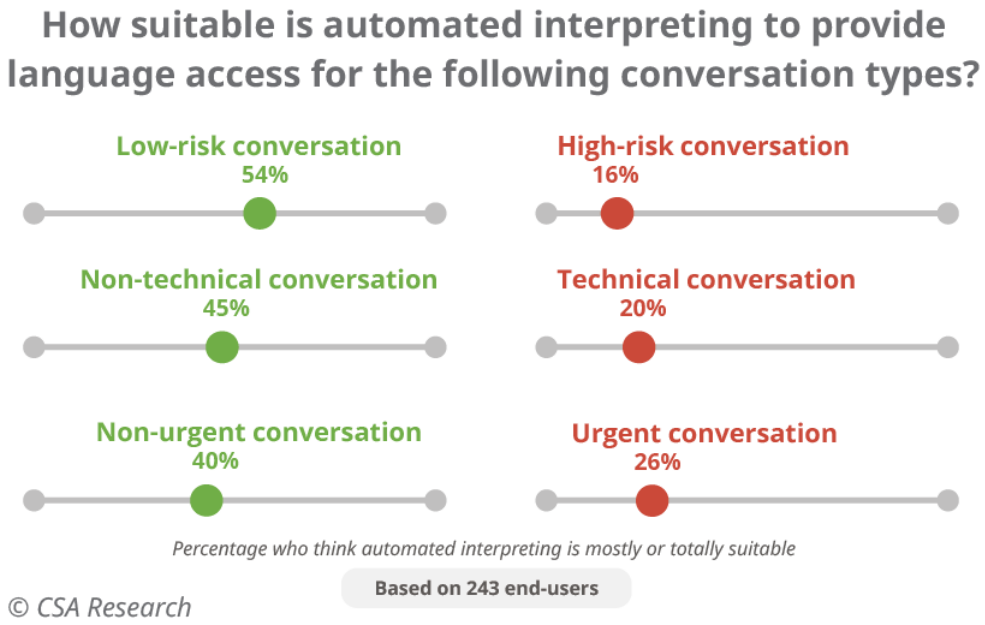
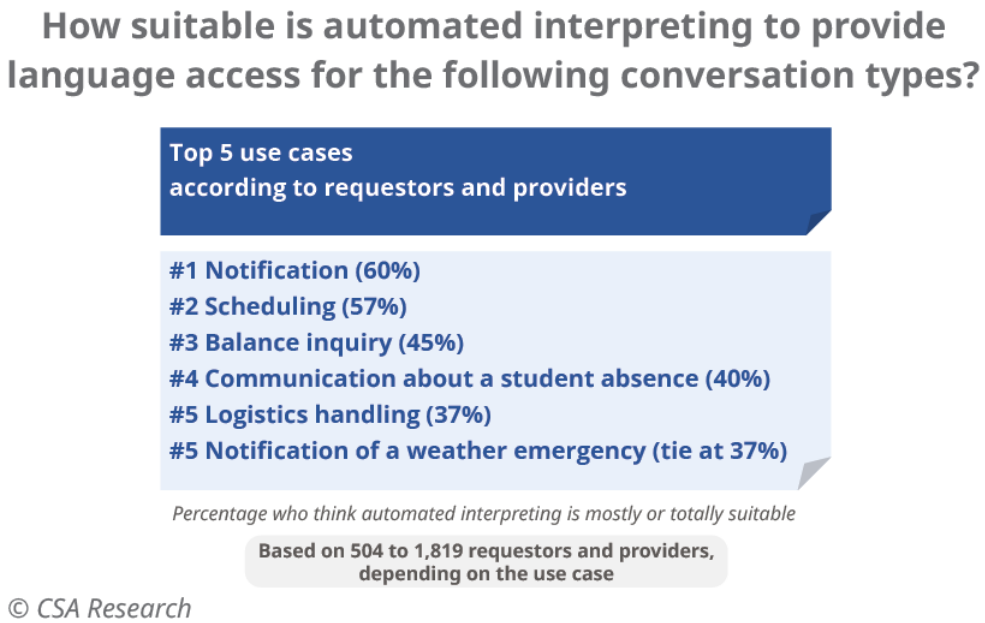


Figure 16: Top Uses Cases According to Requestors and Providers



Respondents overwhelmingly conveyed the idea that, in its current form, AI is not suited for most use cases in high-risk environments like health care and legal. However, those are some of the biggest areas for language and communication access spending. CSA Research recommends that standards, associations, and government bodies developing guidelines work directly with technology vendors to identify and document reasonable use cases. At times, technology vendors have been painted as the enemy, yet they have themselves conducted studies on their own products and can provide guidelines on when they would and would not recommend using them. Critically, guidelines developed without their participation are unlikely to be effective when applied to them (The Challenge Is to Define Suitable Scenarios in Chapter 15).

Figure 17: Interactions Where Automated Interpreting May Be Suitable

Figure 18: Interactions Where Automated Interpreting Is Not Suitable

The Role of the Interpreter

What do all these findings mean?

The traditional view of interpreting revolved around the concept of linguistic conduit, whereby interpreters act as neutral parties who convey messages from one language to another without adding, omitting, or altering the content. This is what automated interpreting also strives to accomplish (Address the Role of Interpreters in Chapter 15).

However, the interpreter’s role has evolved over time to be more than translators of spoken or signed words. They work to verify that the message is understood correctly by both parties, potentially adapting language, tone, or cultural references to fit the context, clarifying meaning, and sometimes even mediating the conversation. This approach recognizes the interpreter as an active participant in the communication process, acknowledging that their presence and decisions can influence the outcome of the interaction. AI is nowhere close to capable of reaching this level of proficiency (Address the Role of Interpreters in Chapter 15).

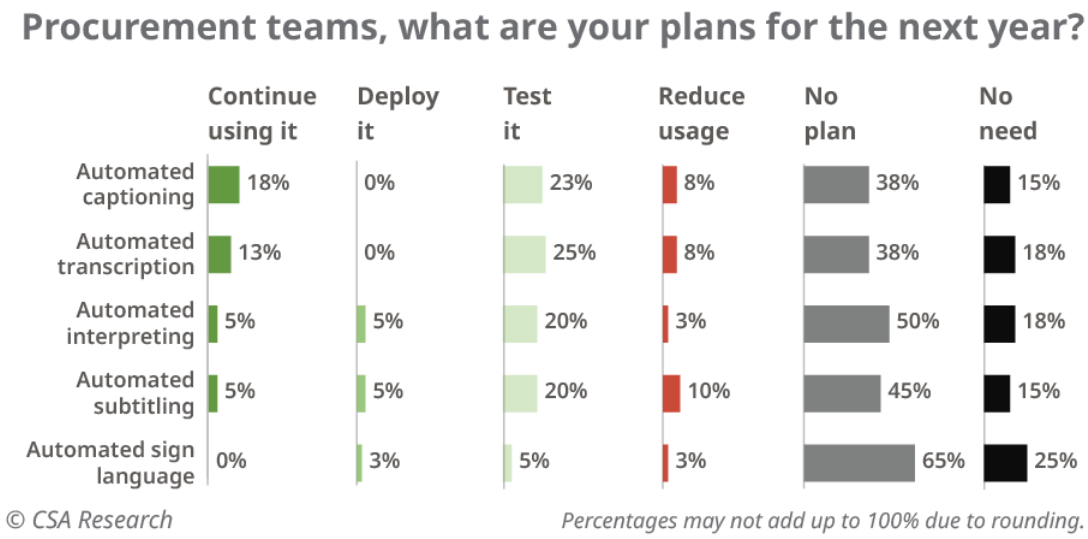
Expecting that not too long from now automated spoken language interpreting will achieve the skills needed for conduit interpreting for some languages in unambiguous, culturally neutral situations is realistic. However, the market expects active interpreting. Organizations and end-users have to define their needs and preferences. And if they choose automated interpreting, how will organizations address elements lost when no interpreter is involved (Address the Role of Interpreters in Chapter 15)?

Interpreters frequently fulfill other roles in addition to relaying meaning. Especially in on-site consecutive scenarios, they may act as a cultural broker/mediator, patient advocate, and supportive ally. They do this to advocate on behalf of the end-user when biases are in place. What would it mean for service recipients if they were to lose their cultural broker and advocate? (No Substitute for Human Skills in Chapter 9)

The bottom line of this study is that lot of the negative perspectives from respondents in this report aren't tied to technological capability – especially as so few tested it – but to how interpreters will continue to make a living if the cost of conduit interpreting in their language becomes null or negligible (The Existential Crisis in Chapter 15).

There is time to resolve this. Procurement teams are not planning mass implementations of AI solutions any time soon (Plans for the Following 12 Months in Chapter 13).

Figure 19: Procurement Teams' Plans for the Next 12 Months



Recommendations

Perceptions on Automated Interpreting

> *"Like most new technologies, AI will be misused in the beginning until consumers learn when and how to use it." [Interpreter in California, no AI experience]*

AI is bringing great changes to many facets of the language industry, from transcription and machine translation to synthetic voice and auto-generated images. Although human interpreting may be preferable, the reality is that it is not always available when people need it. Automated technologies can make a real difference in helping individuals participate fully in society and take advantage of opportunities that they might not otherwise have access to.

Based on our analysis of 2,543 responses to a large-scale study on automated interpreting, CSA Research formulated the following recommendations for the Interpreting SAFE-AI Task Force, requestors, service providers, and technology vendors.

For The Interpreting SAFE-AI Task Force

- **Don't let the stakeholder group's perspective skew the analysis.** The mission of the Task Force is not to decide "whether" to use AI but provide guidance on "when" and "how" to use it. For example, interpreters who fear for their jobs may not be the best judges of what is best for end-users. Various stakeholder groups will be affected in their own ways – each valid and essential to include in the next-step analysis.

- **Continue the research with deeper scenario-based analysis.** Establishing guidelines will require examining accuracy and risks as well as technological capabilities. Follow-up studies must capture the end-user perspective. This does not have to be through a survey. It can be through focus groups or by rating of interpreting examples from different sources. Failure to capture end-user perspectives will lead to linguicism – discrimination based on language.

- **Conduct follow-up research on exclusion areas.** Signed language will benefit from a separate study, as will AI transcription, AI captioning, and AI subtitling processes.
- **Account for lack of end-user feedback.** Absence of data is telling in its own way. The fact that few end-users participated in the current research is not just a recruitment issue. Lack of knowledge or interest in the topic means that the language industry must guide the decision-making process when it comes to guidelines while finding ways to uncover and integrate more end-user feedback.
- **Replicate the same study at periodic intervals.** Technology is evolving fast, and perceptions will evolve as more stakeholders gain experience. Areas where AI currently presents more risks than benefits will change, thereby requiring a regular update of recommendations. Asking the same set of questions one or two years after this original study will enable a point of comparison to assess evolution in acceptance.
- **Investigate which conversations meet which criteria.** Analyze actual conversations to determine what percentage of time they are at an AI-optimal level versus the percentage of time they veer into more challenging territory.
- **Compare results by system and language.** Different AI tools deliver different levels of quality. Even within one tool, different language combinations – or topics – will trigger different quality scores.
- **Educate the market.** Provide training and guidance to the language access ecosystem on "how the automated interpreting magic happens" to contribute to better ability for informed decisions.

For Requestors

- **Assess use cases wisely.** Don't buy into all the hype without understanding the consequences. In many cases, recipients of interpreting services are vulnerable audiences that will lack the ability to advocate for better services. This survey confirms the difficulty of reaching them to document their opinions, often due to

the fear that doing so will affect them negatively. However, you have to do what is right by them, even if they can’t or won’t voice their perspectives.

- **Delineate acceptable use case scenarios.** Automated interpreting is not an all or nothing scenario. The data indicates valid use cases exist. Assess what those are for your organization versus those that should never be handled by automation and those in the middle that depend on specific scenarios.

- **Start with conferencing platforms.** Inquire about built-in capabilities or available integrations. Test specialized tools, which can offer a great way to add multilingual capability to meetings to boost comprehension when participants possess only an imperfect knowledge of the presentation language.

- **Remove the stigma.** Most organizations frown upon the idea of using machines yet already experience them in everyday conversations. For example, when you call a bank or utility company, you probably start with a bot that triages the request and escalates the conversation to a human call agent only when it reaches the limits of its capabilities. There is no longer a need for a human to confirm your credit card balance. Likewise, there is no need for a human interpreter when a nurse checks a patient’s temperature or if an airline attendant processes a basic flight rebooking.

- **Plan for escalation.** Many of the risks from automated interpreting can be mitigated if you create a clear path to bring in human interpreters when needed. Design this capability from the beginning and monitor how people use it. Avoid people getting stuck with inadequate machine services and no way to access help.

- **Brainstorm new use cases.** As with machine translation, which does not replace human translation, the alternative to machine interpreting is seldom human interpreting but rather no interpreting. It is thus useful to fill gaps in language and communication access that would otherwise be unmet. Accordingly, for now, don’t try to insert automated solutions into scenarios where you already use language services – you are likely to be disappointed by machine interpreting’s performance. Instead, inspect where you don’t offer language and communication access, where automation could help triage requests, provide support in understanding, or deliver basic answers.

- **Collect and analyze requirements.** What type of session needs interpreting: live, on-demand, pre-recorded? How frequently? In which languages? Do you expect conduit or active interpreting? Do you have budget limitations? Do you understand the demographics of the end-users?

- **Learn how to select and vet a solution.** Business options are limited today, but due diligence is still required to assess system capabilities. Accuracy and use case applicability vary greatly across solutions. Engage tech-savvy staff to review specifications and validate AI vendor claims, along with people with the right experience in target languages and subject domains to test accuracy for scenarios that replicate your use cases.

- **Pay attention to the terms.** Confirm that prospective tools abide by data privacy standards. Ideally, they should store no more than a few seconds of a conversation and enable offline interpreting to ensure that data isn't used to enhance engine performance.

- **Focus on suitable use cases.** Don't waste resources on applying automated interpreting to areas that it simply can't handle. For example, many people hope for relief through automation to better support languages of limited diffusion. However, the lack of appropriate training data makes it one of the worst use cases for the foreseeable future. Instead, focus on one-way communications and templated basic conversations.

For LSPs and Interpreters

- **Add extra value.** AI can only provide value equivalent to interpreting words. Make your clients aware of how frequently you do more than "just" interpret words. Your cultural advisor skills will be what sets you apart from AI output.

- **Gain first-hand experience with the tools.** Only 9% of interpreters extensively or moderately tested AI interpreting. Know your competition. You are no longer just competing against bilingual non-professionals and remote interpreters. You need first-hand experience to ground your differentiation in facts. And when testing tools, put your own bias aside. Don't look at it from the fear of losing your job but for what it can do for the constituencies you serve – position yourself as an expert in suitable

and unsuitable use cases. Many responses in the survey showed angst without concrete knowledge of the tools.

- **Experiment with AI to augment your interpreting.** AI does not need to just be about replacing human interpreters. It has the capabilities of giving extra input to professionals when using it in a computer-aided interpreting (CAI) solution. It can help mine reference documents for terminology to prepare ahead of a session and display source and targets on screen when they are brought up in the conversation – making the recall of new terms easier. It is also helpful to display on screen details such as numbers, people, company, or product names to reduce the recollection burden.

- **Work with technology vendors to reduce the cognitive load.** CAI solutions add yet another stream of information for interpreters' brains already busy capturing what was said, understanding it in the context of the session and the visual cues available, and rendering it in a different language. Collaboration between tech vendors and interpreters is essential to ensure CAI tools help and don't overload interpreters.

For Technology Providers

- **Build CAI tools.** Integrate them to your telephone, video remote, and remote simultaneous platforms (OPI, VRI, and RSI, respectively). For escalation scenarios, produce a list of important terms already used in the conversation and present suggested translations to the interpreter who's coming into the conversation cold.

- **Improve accuracy, privacy, recognition of limits, etc.** It's okay to shoot for "conduit-level" accuracy first – but deliver on that well. Carefully examine the findings in the What It Would Take to Increase Trust in AI section in the main report as it provides many great pointers on what to work on next. Make sure you also focus on refining tools' "hearing" capabilities and testing them in real-world situations, such as with background noise, low bandwidth, or poor lighting. Today, AI only deals well with good audio and well-built and well-enunciated sentences. A human will deal with imperfect audio or pronunciation better than artificial intelligence.

- **Consider offering different modes.** AI has a single mode. Humans can condense, correct, or simplify information on the fly because they understand what matters and can help the presenter get their message across. Create settings to address common issues and aid understanding – such as adjusting the AI machine to produce the target language at an eighth grade reading level regardless of the register of the source language.

- **Provide testing opportunities.** Too many people have never even experienced some of the more advanced products on the market. Exposure to the technology increases buy-in and reduces the stigma that technology vendors face when selling an automation solution.

- **Invest in and support controlled testing.** Engage third-party testing organizations to conduct independent side-by-side comparisons between human and machine interpretation outputs and between AI interpreting systems.

- **Share your testing data and be transparent.** The market is eager to see the performance of your systems by scenario or language or whatever else you track. Lack of data leads to mistrust. Be open about how you test your systems rather than just presenting numbers: Customers are naturally wary of numbers without context and too-good-to-be-true claims. Exposing how you arrive at results can help allay suspicion.

- **State your goals publicly.** Many respondents fear that AI vendors are out there to replace humans. Help educate industry stakeholders on the targets you pursue and suitable use cases. Failure to spell this out will mean they think you’re going after everything. The industry wants technology vendors to act responsibly.

- **Offer escalation paths.** When you call your bank or utility company, you usually start with an automated system. However, when you reach the limits of what the system can deliver, you can ask to talk to an operator. Likewise, when machine interpreting fails, enable easy escalation mechanisms so a professional can address miscommunication issues.