SAFEAI

Update: SAFE-AI Task Force

May 17, 2024

Federal Interagency Language Roundtable

Interpreting SAFE-AI Task Force
Stakeholders Advocating for Fair and Ethical AI in Interpreting

ChatGPT

- Launched on November 30, 2022 and **crossed 1 million users in just 5 days** of launch and **gained 100 million active users by January 2023**.
- As of May 2024, the website **sees nearly 1.8 billion visitors per month.**

SAFEAI

Concerns

- Privacy (data security)
- Accuracy and Reliability (hallucinations, faulty logic)
- Source data limitations (leads to bias)
- Lack of contextual information (cultural cues)
- Accountability (human in the loop)
- Misuse

Initial Launch Group Members

Katharine Allen, Training Specialist, Boostlingo

Carla Fogaren, RN, Vice President, National Council on Interpreting in Health Care

Cody Francisco, CDI, Director, Deaf & Hard-of-Hearing Services, MasterWord Services, Inc.

Ludmila Golovine, President & CEO, MasterWord Services, Inc.

Winnie Heh, Career Advisor, Middlebury Institute of International Studies

Eliana Lobo, CoreCHI-P, Director, Lobo Language Access

Alan Melby, Chair of FIT North America

Natalya Mytareva, Executive Director, Certification Commission for Healthcare Interpreters

Barry Olsen, Principal Consultant, What about language?

Hélène Pielmeier, Senior Analyst at CSA Research

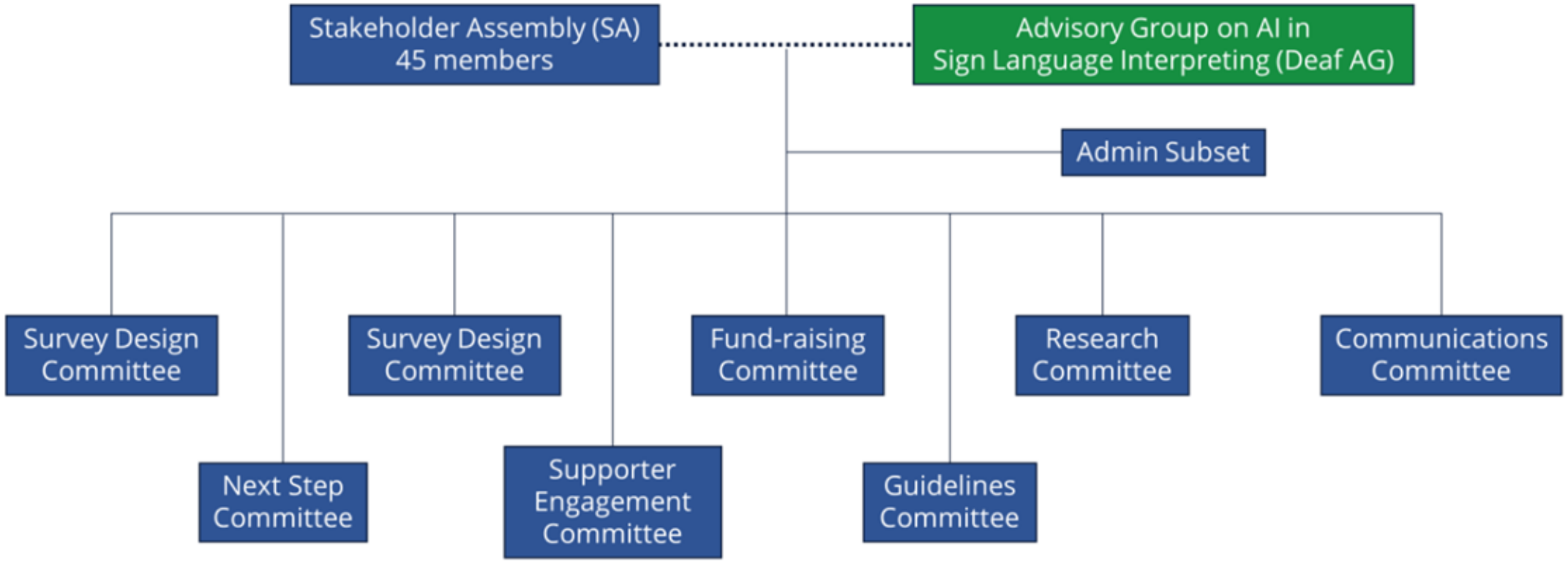
Dr. Bill Rivers, Principal, WP Rivers & Associates

SAFE-AI

Response to Call-to-Action Meeting

- **Anonymous Survey**
- **600** responded to July survey (80% US-based)
- **200+** volunteered for various roles
- **11** Stakeholder groups identified
- **45** member Stakeholder Assembly formed

Governance Structure



Milestones

- Workgroup Launch (6/27/23)
- Formation of Stakeholder Assembly (9/14/23)
- Launched Perception Survey & Deaf AG Study (Nov 2023)
- Published Perception Survey & Deaf AG Report (March 2024)
- Draft Guidelines posted for comment (May 2024)

SAFEAI

Perception Survey: Use Cases for AI in Interpreting
Qualitative Study by Deaf Advisory Group

1. **Perception Survey:**
   - **End-User Survey:** 'Automated Interpreting – Perceptions from Actual Users of Interpreting Services.'
   - **Market Survey:** 'Automated Interpreting – Perceptions from the Market.'
   - All analyses weighted by stakeholder class/group
2. Qualitative Study by Deaf Advisory Group
3. All results are available: **https://safeaitf.org/**

Overview of Findings from “Perceptions on Automated Interpreting”

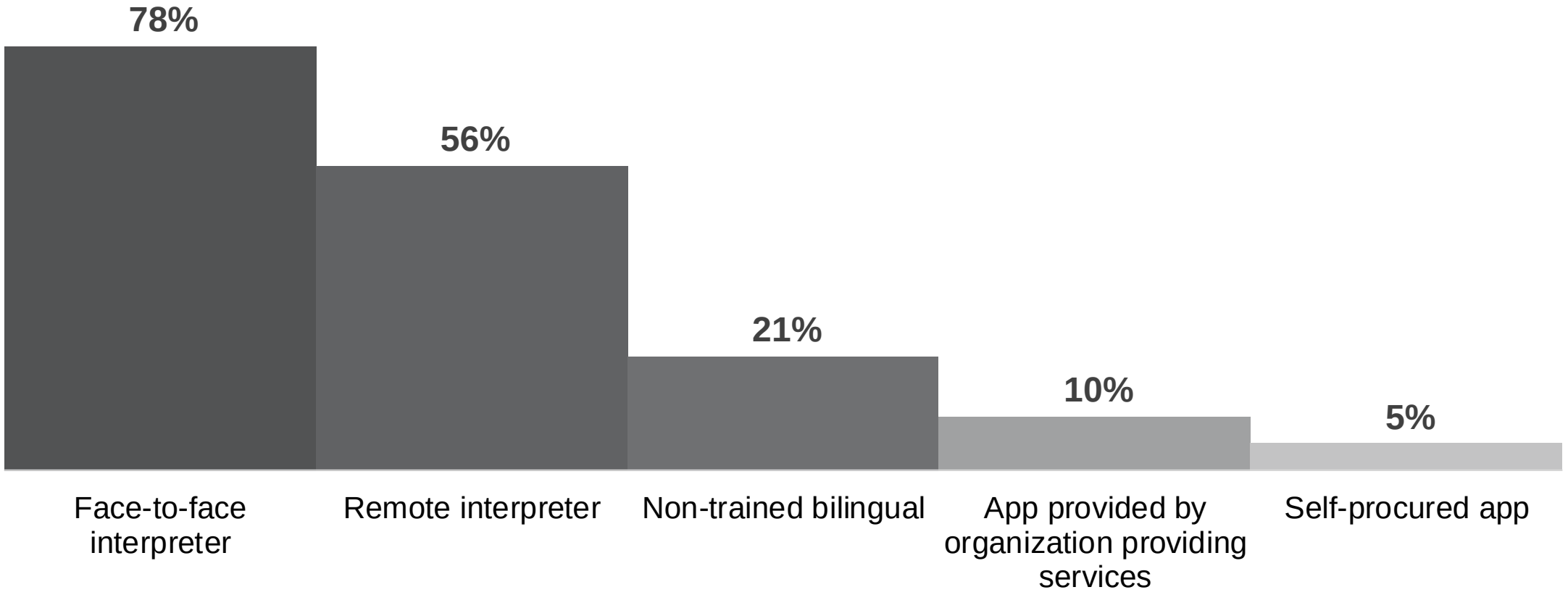
Research by the Numbers

2	surveys
10	languages
118	datapoints collected
2,543	respondents
82	countries
48	US states
3,400	free-text comments
9,400	datapoints published
67%	interpreters
69%	health care
11%	hands-on experience
79%	based in the US

Based on "Perceptions on Automated Interpreting" © CSA Research

SAFEAI

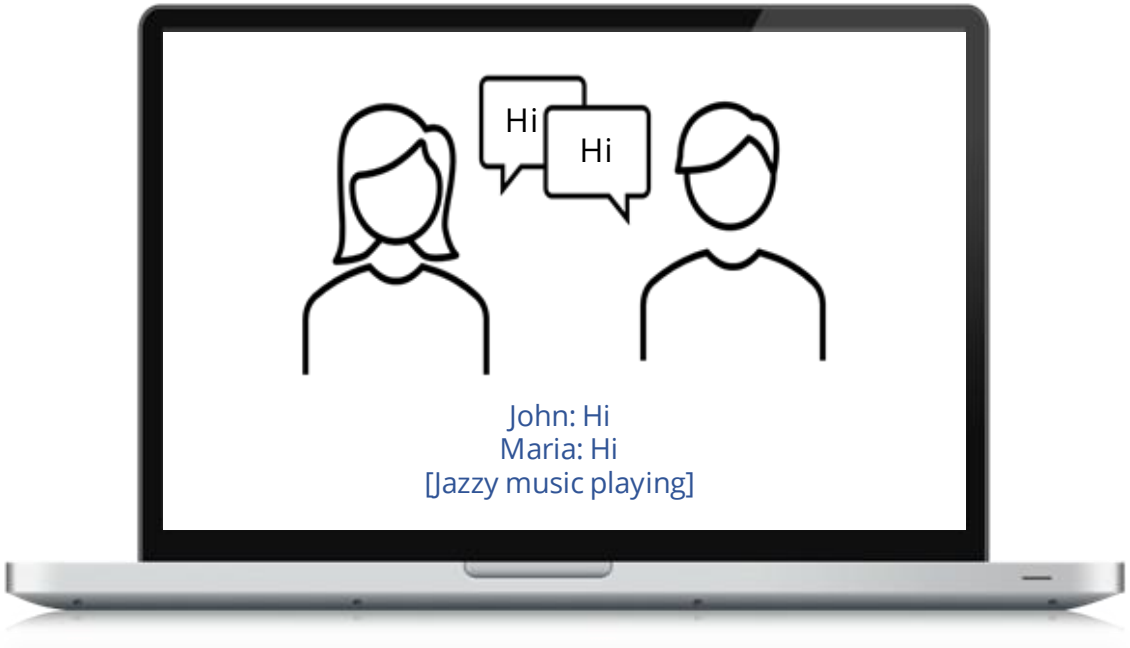
I Fully Trust Interpreting from…

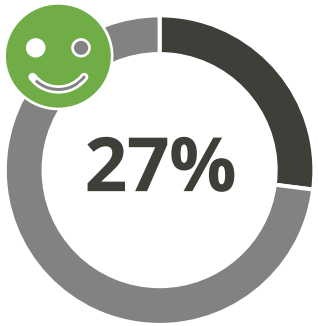


Based on 263 end-users of interpreting services

Based on "Perceptions on Automated Interpreting" © CSA Research

Automated Captioning





of respondents used or tested it moderately or extensively

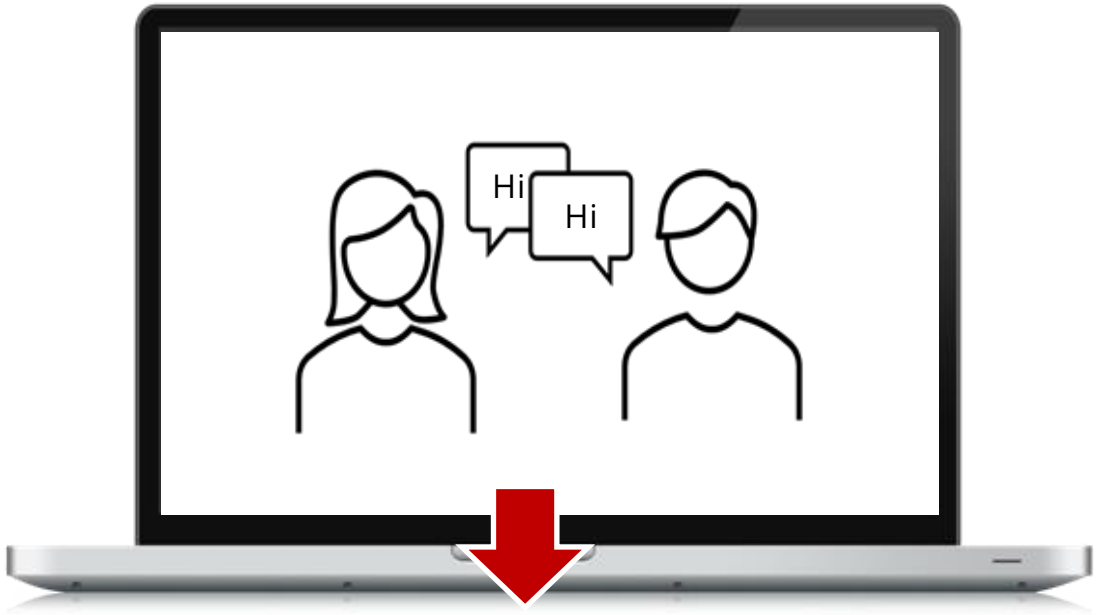
Based on 2,406 answers

of actual users find results good or excellent

Based on 944 answers

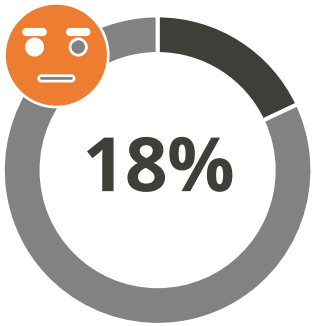
Based on "Perceptions on Automated Interpreting" © CSA Research

Automated Transcription



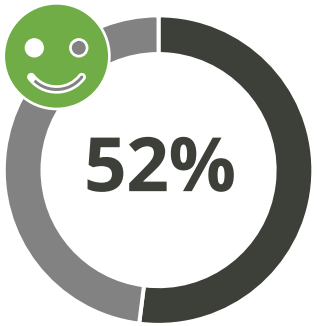

```
100:00:34,000 -->
00:00:40,260
[Maria] Hi

2
00:00:40,760 -->
00:00:44,772
[John] Hi
```



of respondents used or tested it moderately or extensively

Based on 2,157 answers



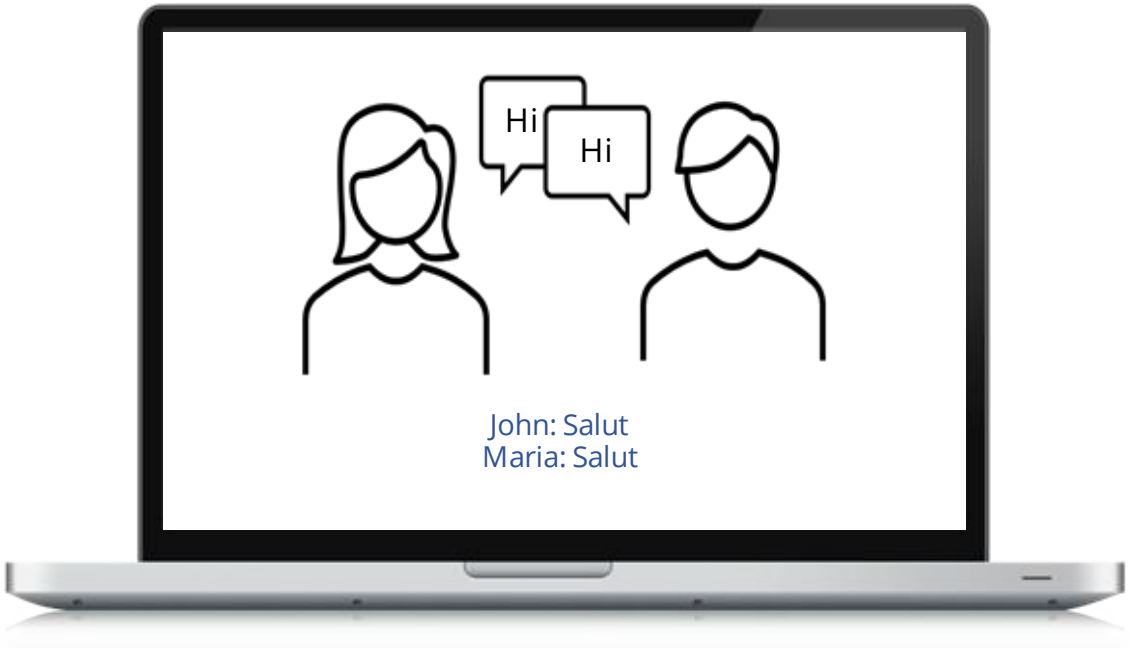
of actual users find results good or excellent

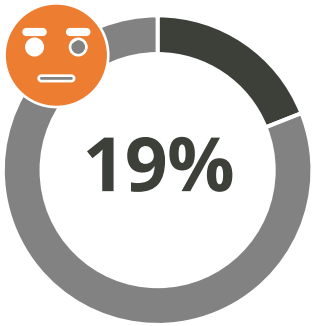
Based on 711 answers

Based on "Perceptions on Automated Interpreting" © CSA Research

SAFEAI

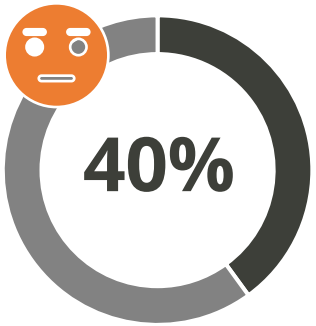
Automated Subtitling





of respondents used or tested it moderately or extensively

Based on 2,157 answers

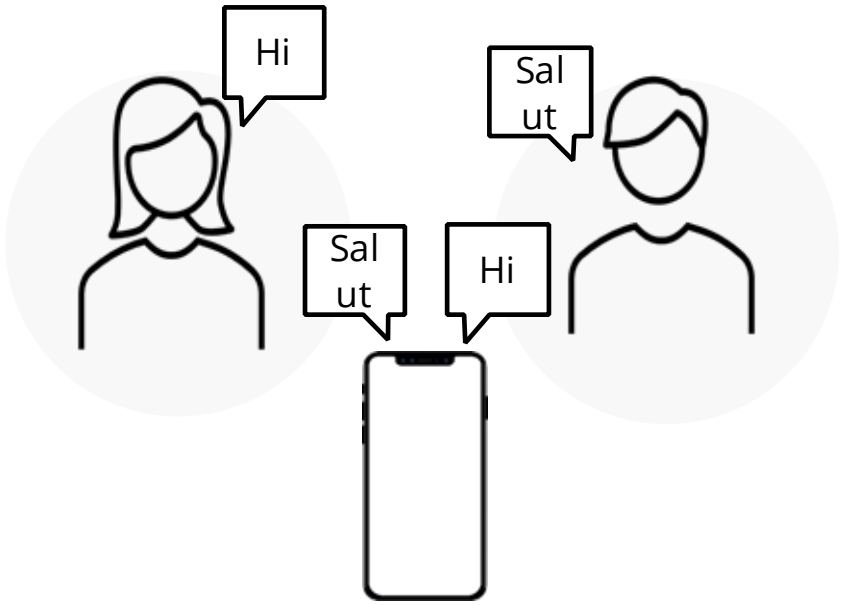


of actual users find results good or excellent

Based on 705 answers

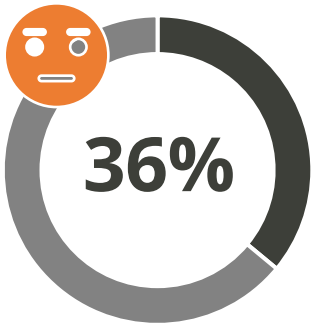
Based on "Perceptions on Automated Interpreting" © CSA Research

Automated Interpreting



of respondents used or tested it moderately or extensively

Based on 2,406 answers

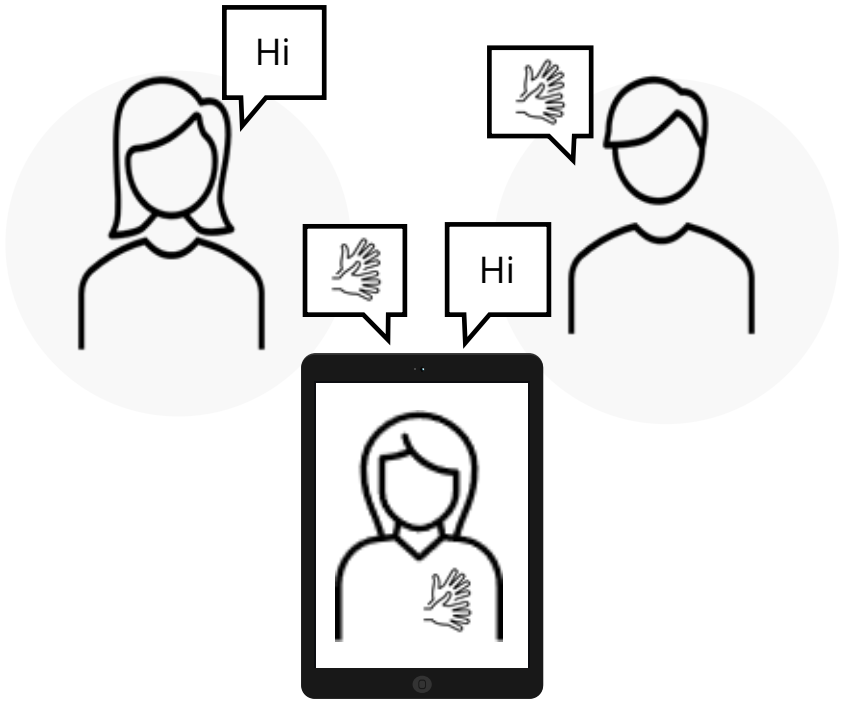


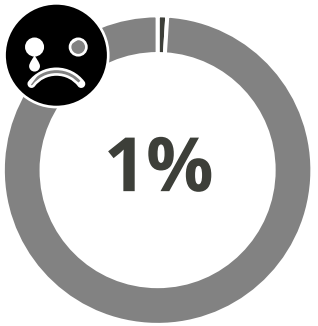
of actual users find results good or excellent

Based on 625 answers

Based on "Perceptions on Automated Interpreting" © CSA Research

Automated Sign Language





of respondents used or tested it moderately or extensively

Based on 2,236 answers

of actual users find results good or excellent

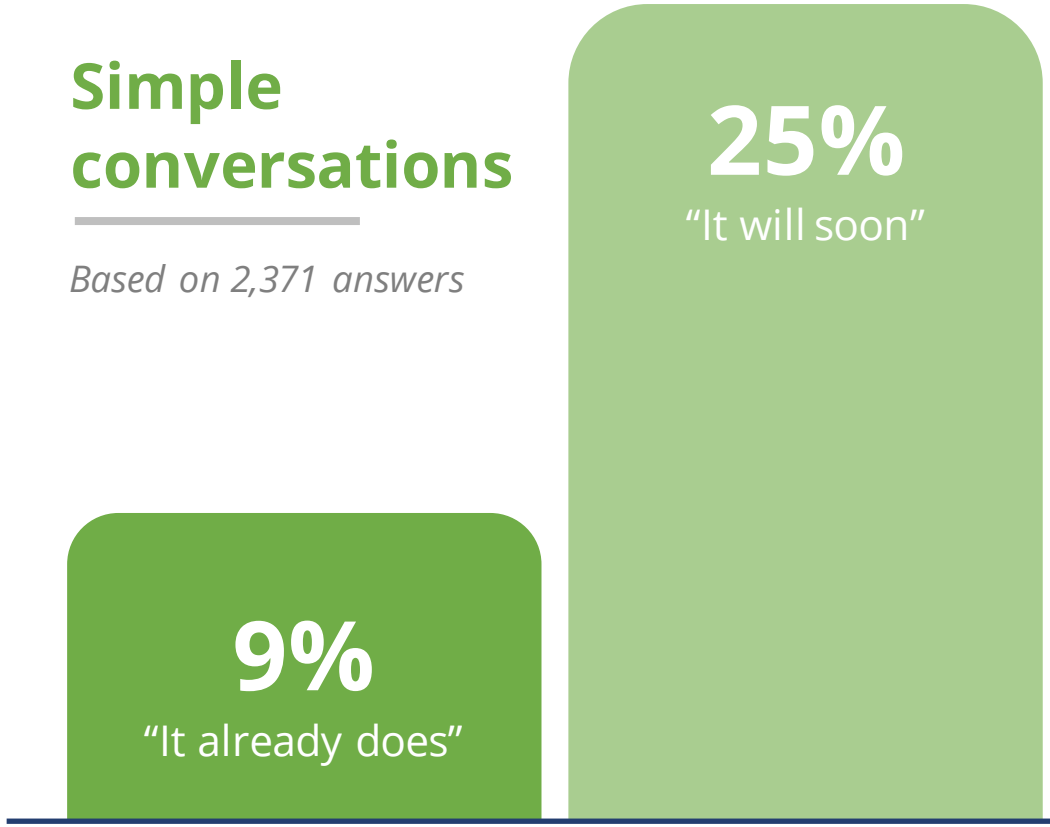
Based on 77 answers

Based on "Perceptions on Automated Interpreting" © CSA Research

Do you think that automated interpreting can reach the same level of accuracy as qualified human interpreters?

Simple conversations

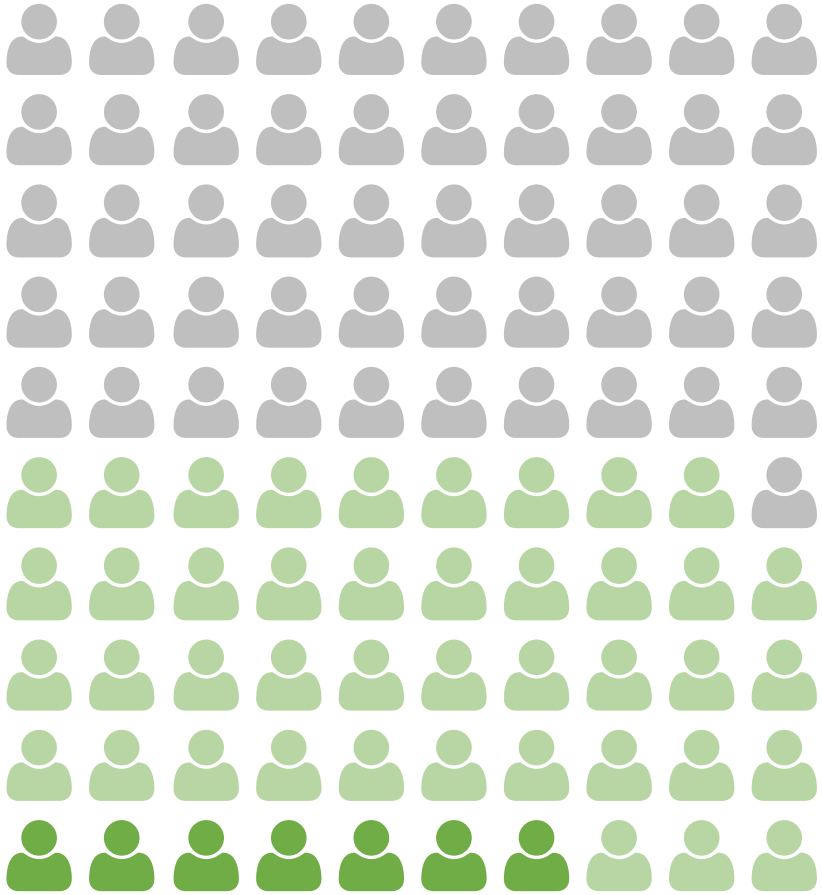
Based on 2,371 answers



Complex conversations

Based on 2,326 answers

Based on "Perceptions on Automated Interpreting" © CSA Research



"It's okay to use machines for routine and repetitive conversations that they can handle"

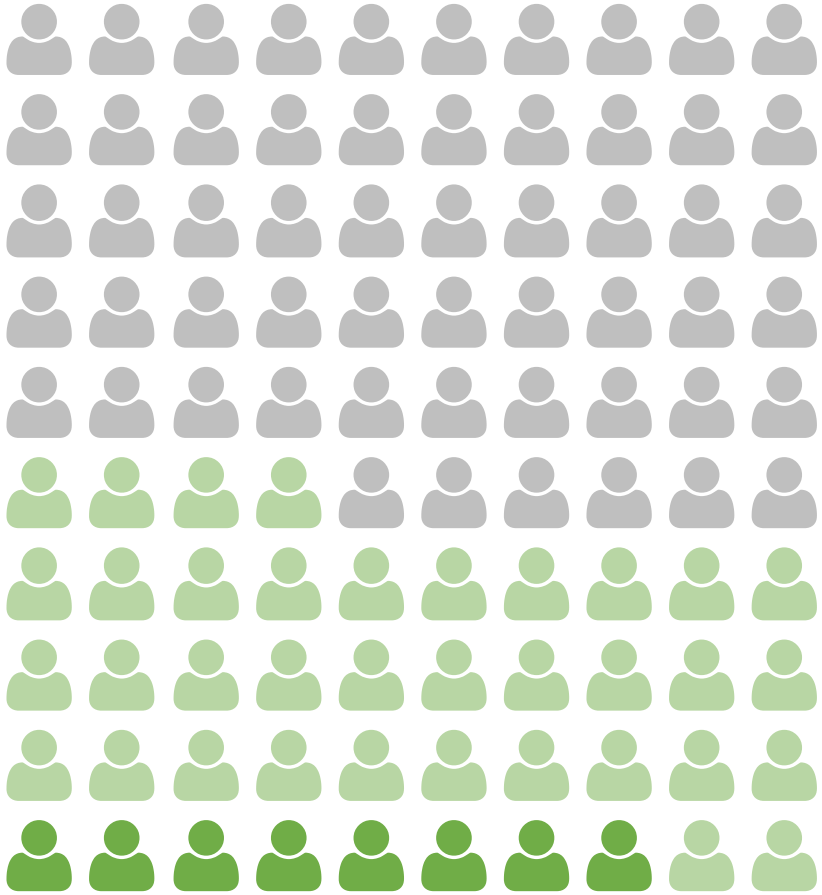
49%

Based on 2,256 answers

7%	42%
Strongly agree	Agree

Based on "Perceptions on Automated Interpreting" © CSA Research

SAFEAI



"Having a machine interpret is better than having no interpreter"

44%

Based on 2,256 answers

8% Strongly agree

36% Agree

Based on "Perceptions on Automated Interpreting" © CSA Research

SAFEAI

"It's not right to replace people with machines for interpreting"

74%

Based on 2,256 answers

48% Strongly agree

26% Agree

Based on "Perceptions on Automated Interpreting" © CSA Research

"Clients only want automation to reduce costs"

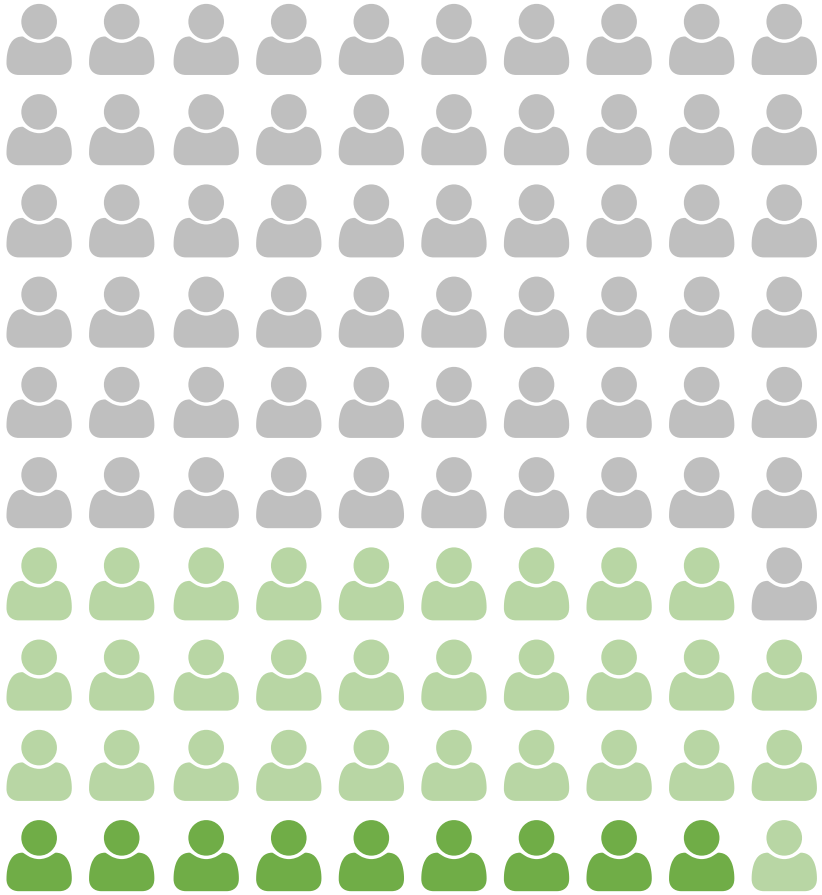
68%

Based on 2,006 requestors and providers

35% Strongly agree

33% Agree

Based on "Perceptions on Automated Interpreting" © CSA Research



“If I have to pay for the interpreter myself, I might choose a machine to interpret”

39%

Based on 250 end-users

9% Strongly agree	**30%** Agree

Based on “Perceptions on Automated Interpreting” © CSA Research

SAFEAI

"If someone else is paying, I might choose a person to interpret"

76%

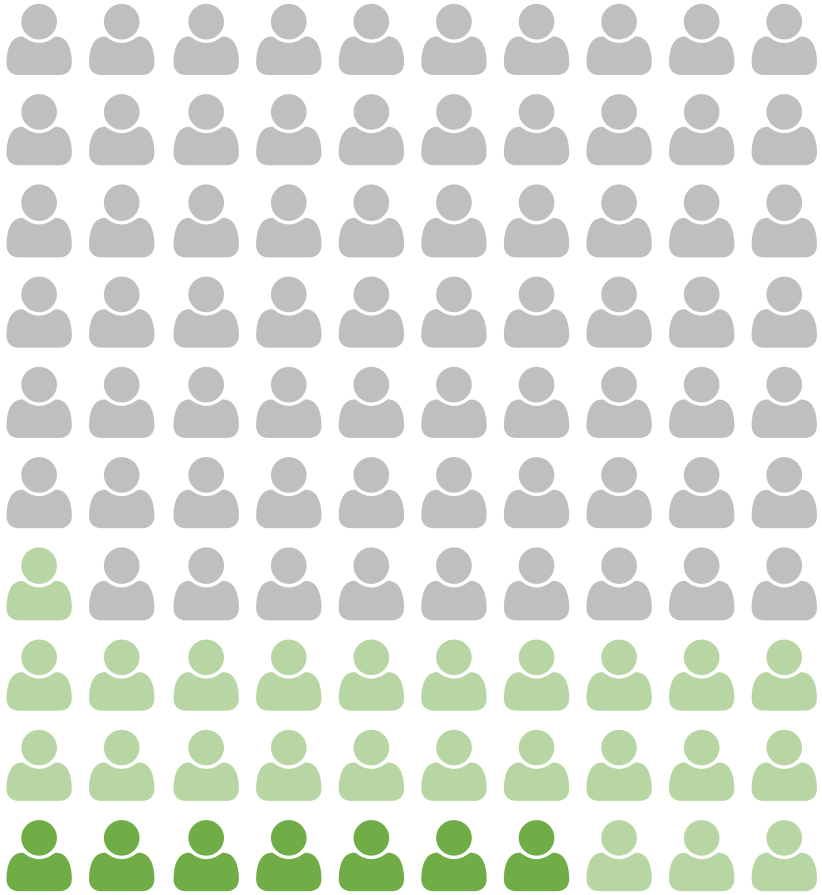
Based on 250 end-users

46% Strongly agree

30% Agree

Based on "Perceptions on Automated Interpreting" © CSA Research

SAFEAI



"There are situations when I would prefer automated interpreting over a human"

31%

Based on 250 end-users

7% Strongly agree

24% Agree

Based on "Perceptions on Automated Interpreting" © CSA Research

"Using a machine interpreter causes me more stress"

51%

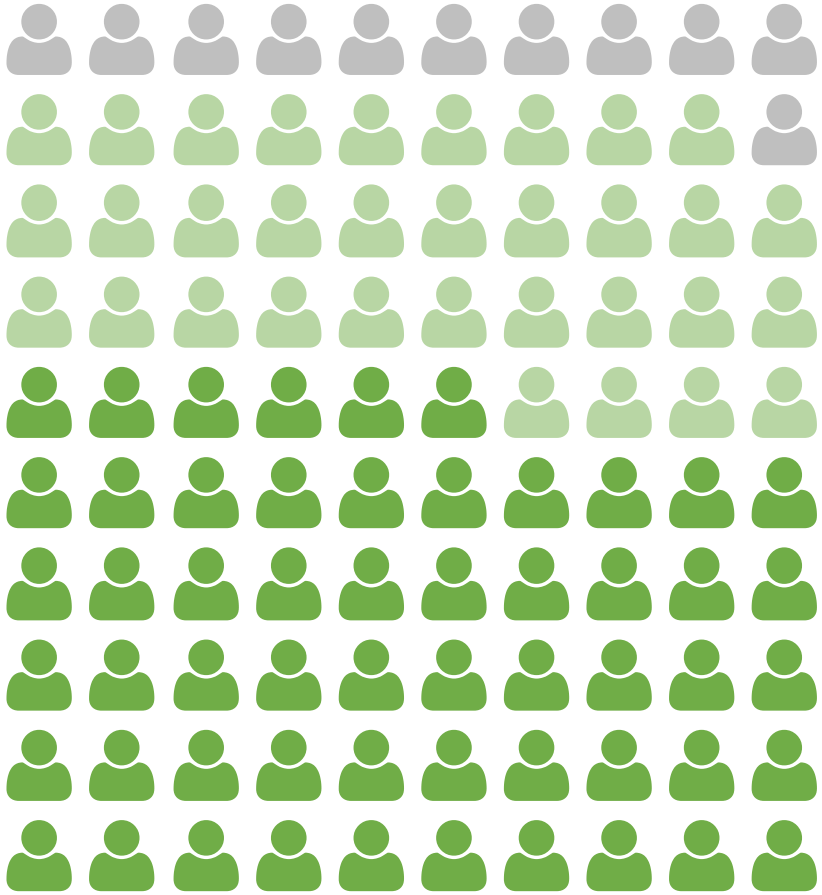
Based on 250 end-users

25% Strongly agree

26% Agree

Based on "Perceptions on Automated Interpreting" © CSA Research

SAFEAI



“I want to know if a person or machine is doing the interpreting”

89%

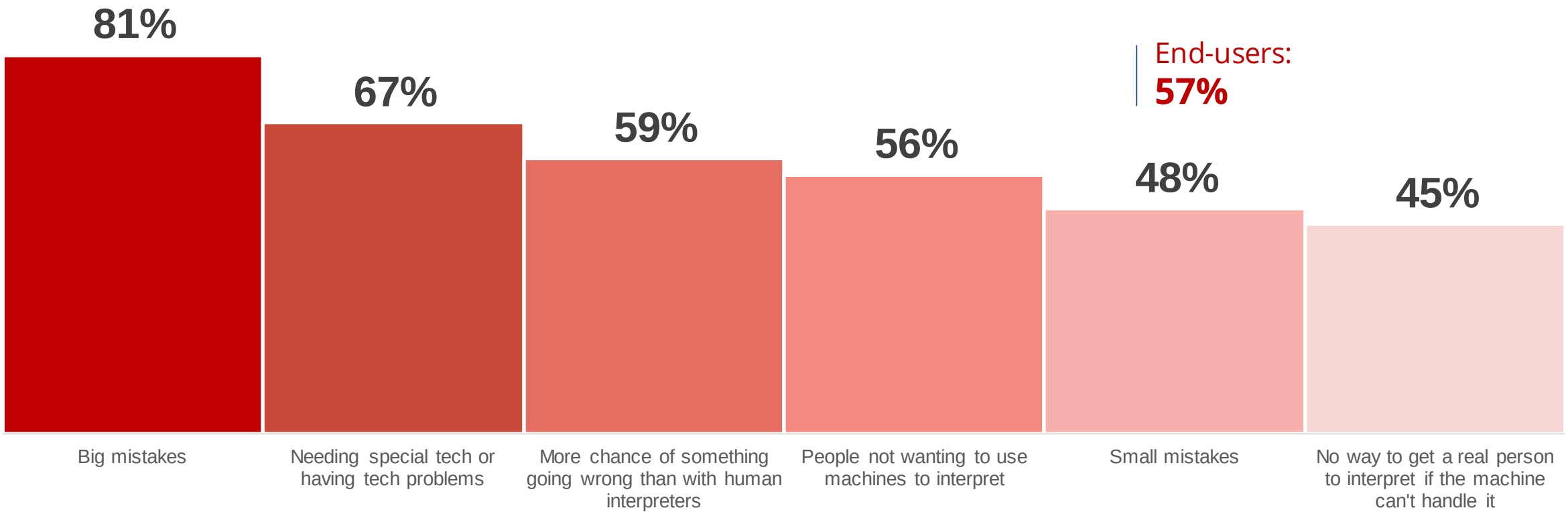
Based on 250 end-users

56%
Strongly agree

33%
Agree

Based on “Perceptions on Automated Interpreting” © CSA Research

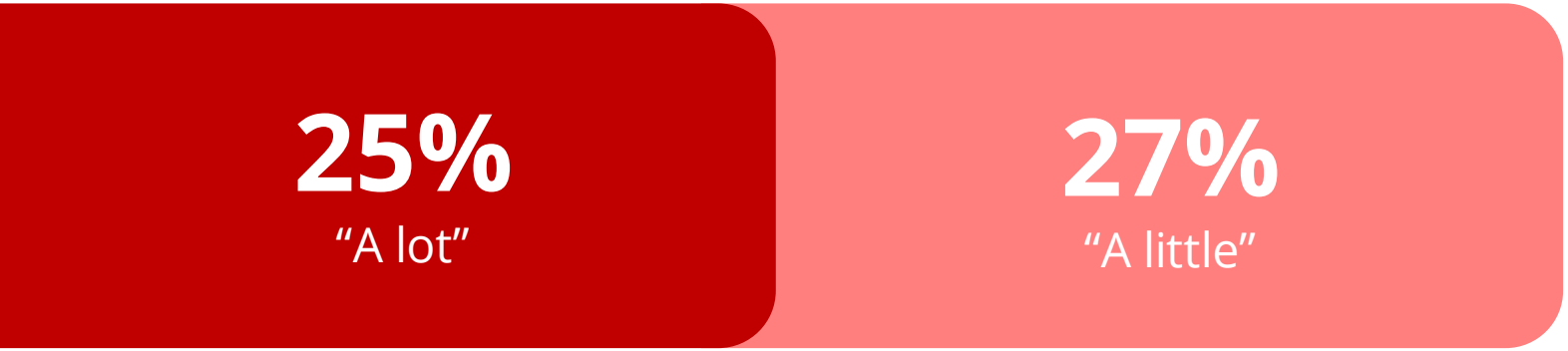
What do you see – or expect to see – as the drawbacks of automated interpreting?



Based on 2,256 answers

Based on "Perceptions on Automated Interpreting" © CSA Research

If a solution can quickly get a person to help when there's a problem, are automated solutions more helpful?



Based on 1,729 answers from requestors and providers

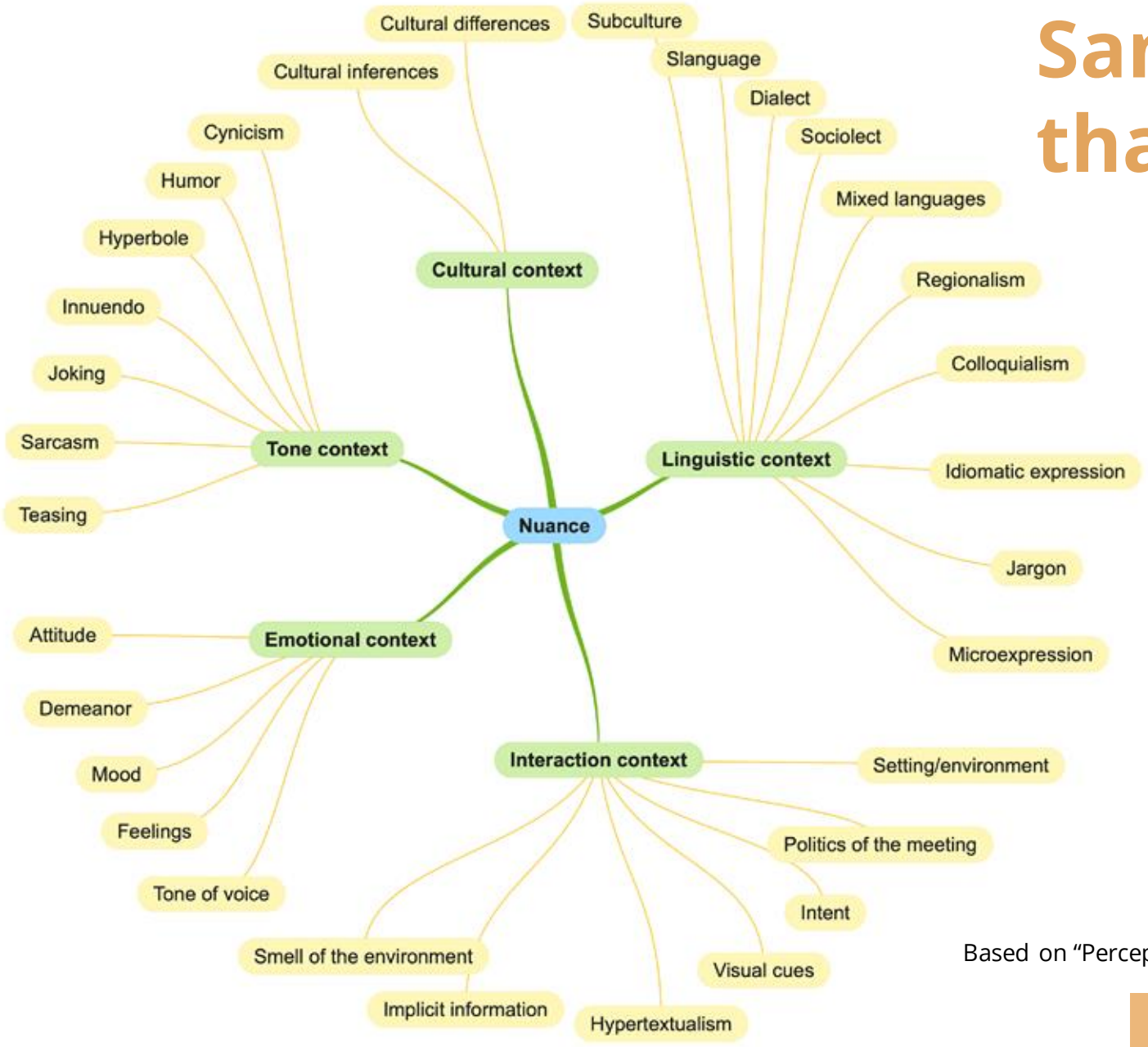
> ***"The best use case would be to have an automated interpreting solution to instantly attend to an emergency while simultaneously working to locate and subsequently relay the call to a real person. Both could then report back to the requester independently, creating a feedback mechanism that benefits AI training and the accuracy of the human interpreter."***

[Interpreter in New York, no AI experience]

Based on "Perceptions on Automated Interpreting" © CSA Research

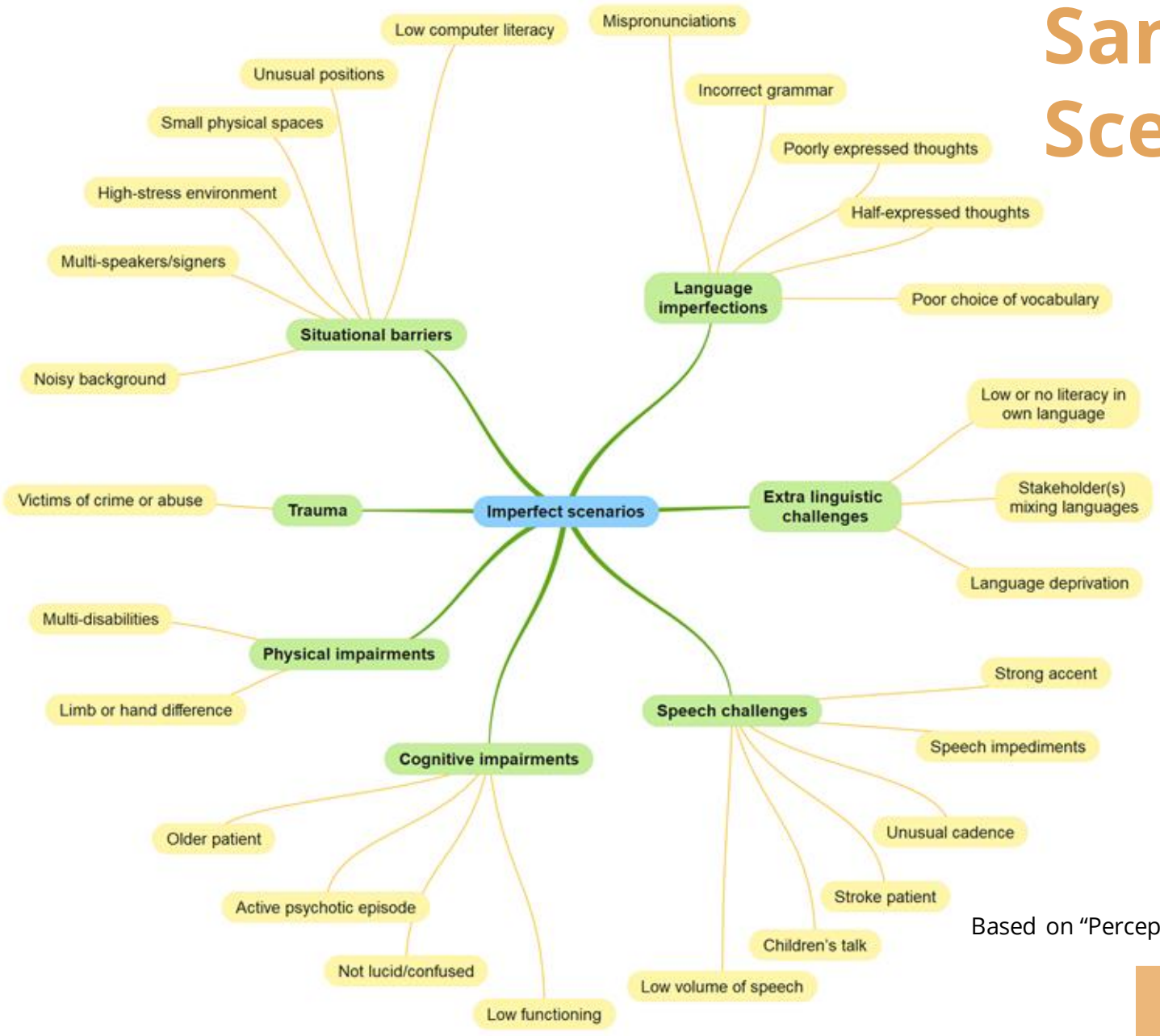
What Trips Up AI

Sample Elements that Require Nuance



Based on "Perceptions on Automated Interpreting" © CSA Research

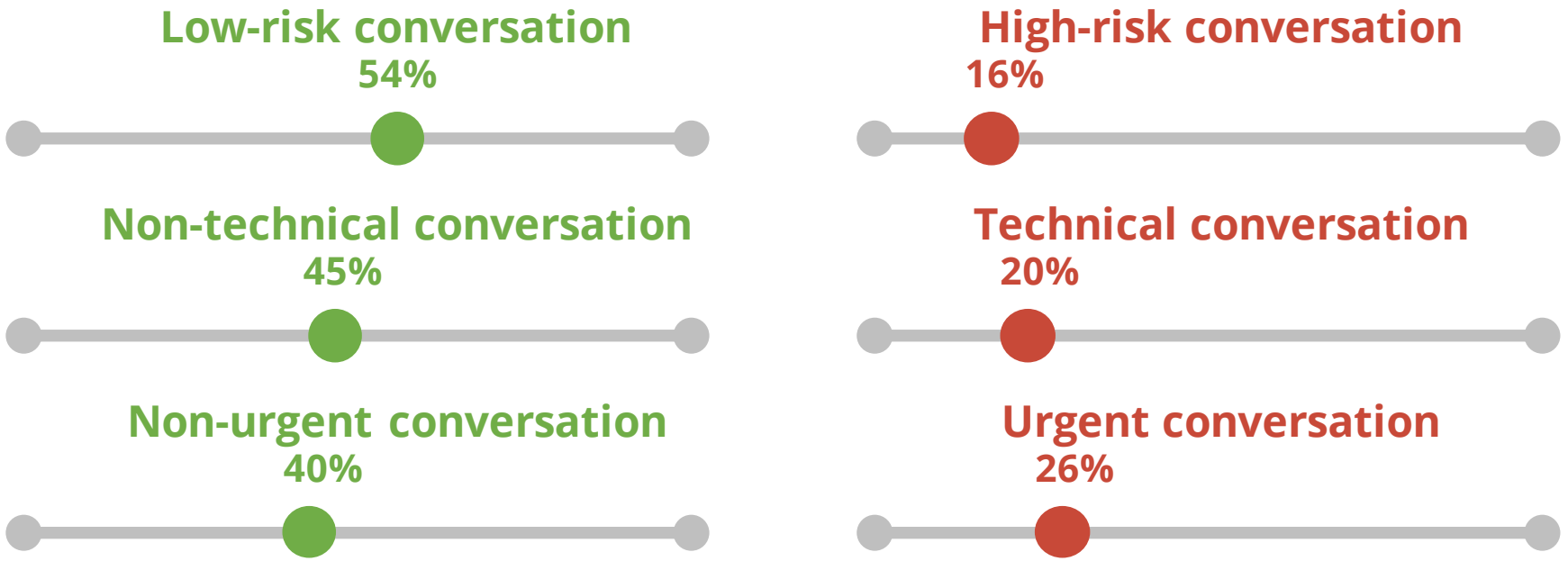
Sample Imperfect Scenarios



Based on "Perceptions on Automated Interpreting" © CSA Research

Suitability by Use Case Scenario

How suitable is automated interpreting to provide language access for the following conversation types?



Percentage who think automated interpreting is mostly or totally suitable | *Based on 243 end-users*

Based on "Perceptions on Automated Interpreting" © CSA Research

Some Points to Consider

1. Beginning snapshot of perception *before* adoption
2. Adoption is accelerating
    - Highest adoption: conference and business interpreting
    - Universal translator syndrome
    - Critical areas: regulated markets
3. Increasing legislation and guardrail guidance
4. Inflection point level disruption to interpreting

ADVISORY GROUP
on AI and Sign Language Interpreting
Report to the Interpreting SAFE AI Task Force Stakeholders Assembly

#DeafSafeAI

Deaf-Safe AI:
Toward a Legal Foundation for Ubiquitous Automatic Interpreting

#DeafSafeAI

What the Study Finds

- Three critical impact areas
- Opportunity to build on deaf lived experience of harms
- Value of a Big Picture lens on possibilities

Ethics and Fairness Methodology

Findings

- Desired Results and Outcomes
- Quality = Accountable Technology
- Readiness depends upon Ethics

Discussion: Risk and The Future Recommendations

Ethics and Fairness

The ethics of AI emphasizes two interwoven concerns:

- controlling bias and assigning responsibility so that generative models are **effectively designed to prevent harms**
- making transparent, verifiable and unambiguous the **accountability** of decision-makers about the design, development, deployment, operations, evaluation, repair, and continuous improvement of large language models (and other generative models)

Ethics & Fairness

Fairness (whether it is achieved or not):

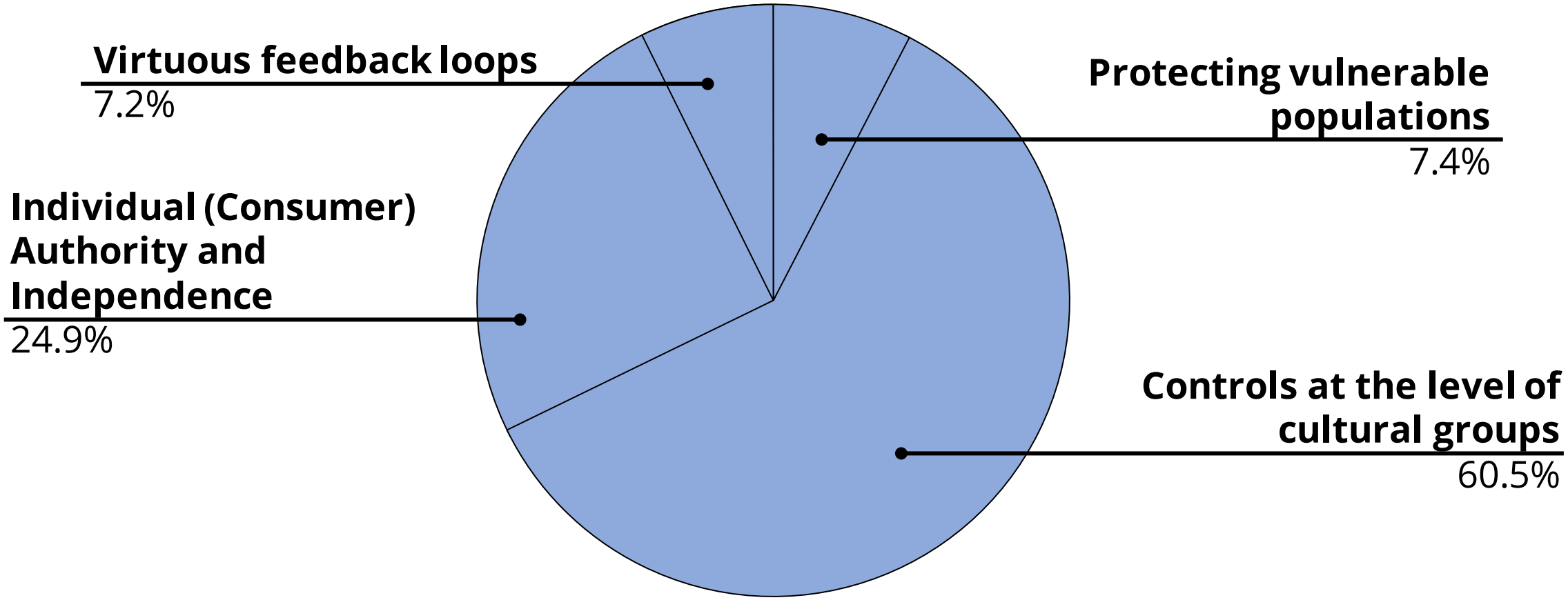
- involves the goal of removing bias/favoritism
- is measured by results, not intentions
- is observable in individual and interpersonal behavior during interaction, and in aggregate statistics → there is material evidence

Key Term

AIxAI

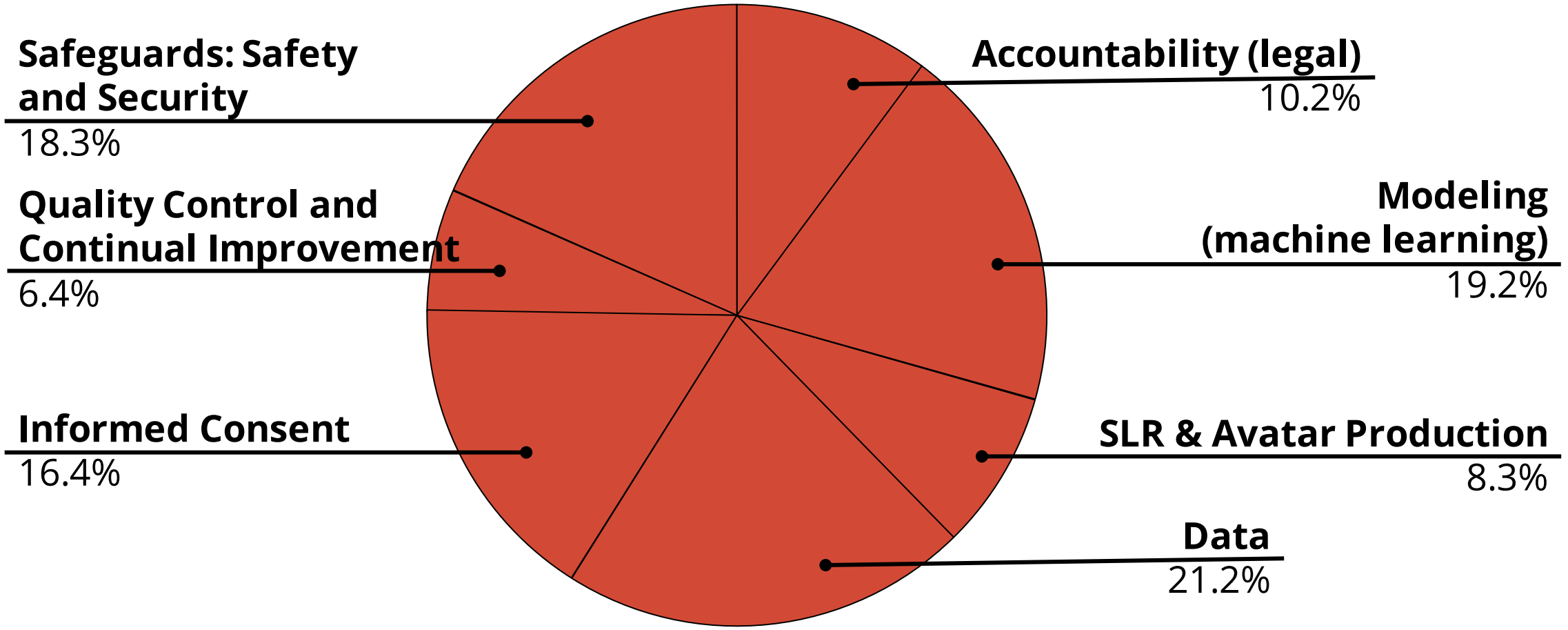
Automated/Automatic Interpreting (AI)
by Artificial Intelligence (AI)

Desired Results and Outcomes



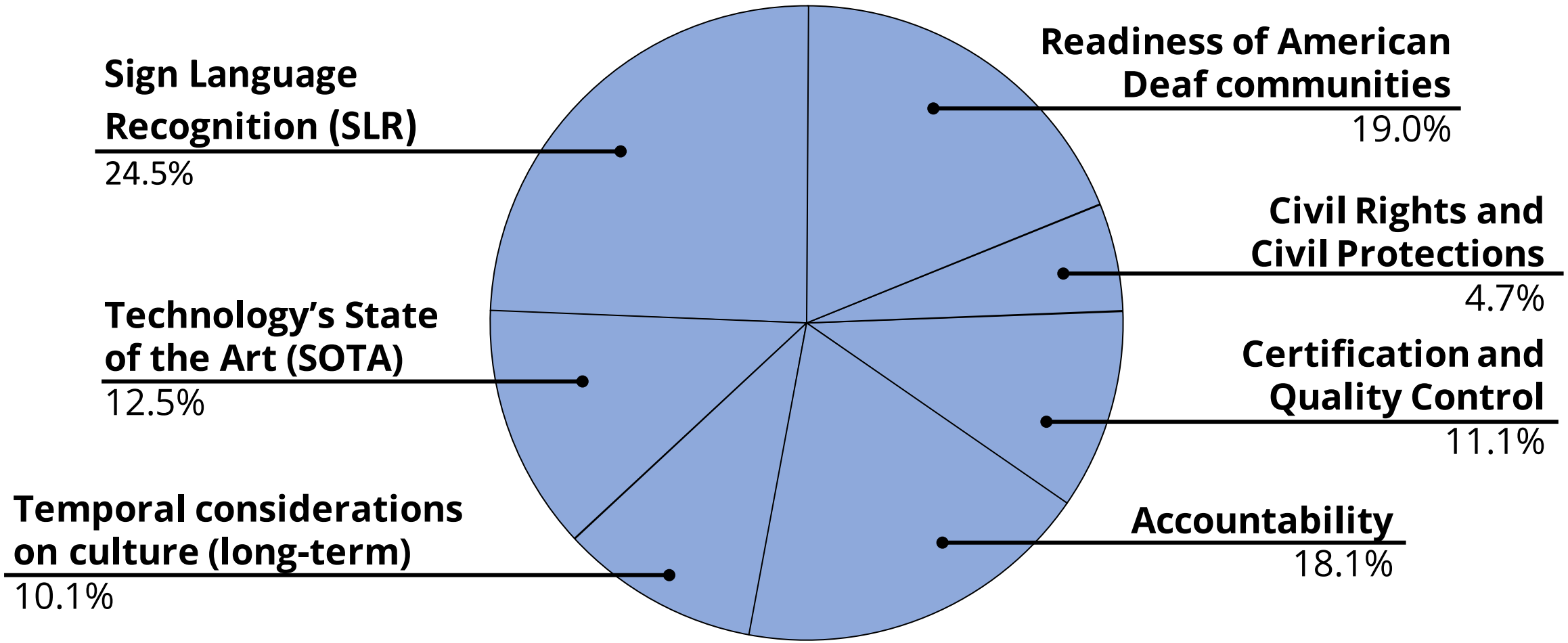
#1 topic of talk = the criteria for controls by designated authorities of the language community

Quality via Accountable Technology



#1 topic of talk = data

Readiness



#1 topic of talk = Sign Language Recognition (SLR)

Risk and the Future

We suggest that

- proper protection of privacy and confidentiality in AIxAI can lead to better protections across the Internet in all kinds of applications
- risks of not putting strict regulations in place will have adverse downstream consequences for governance, business, and social infrastructures

Key Term

Sociotechnical Systems (STS)

Key Term

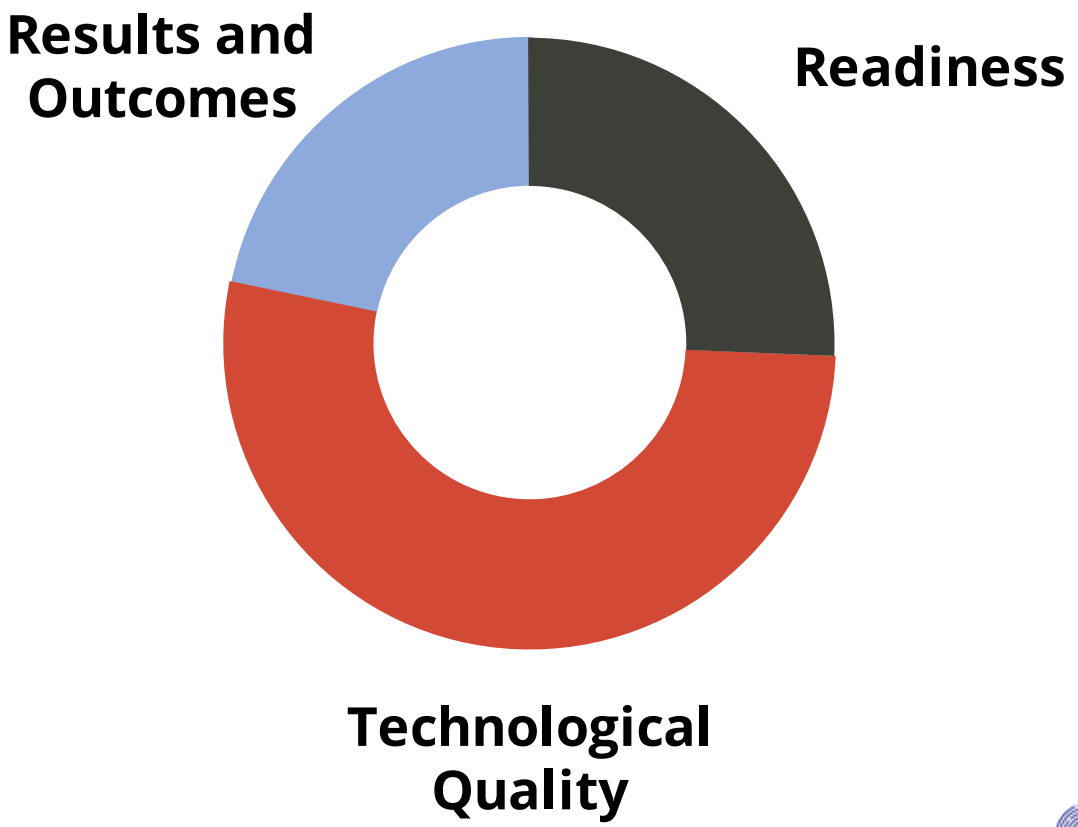
"Sociotechnical refers to the **interrelatedness** of social and technical aspects of an organization. The cornerstone of the sociotechnical approach is the design process that leads to optimization of the two subsystems"

(Botla and Kondur, 2018, p. 26)

Key Term

The KEY is

attending to how the social behaviors of humans combine with, influence, and are shaped by the structures of technology, and vice versa.



Recommendations

- Understand AIxAI as a **sociotechnical system**
- Utilize Deaf wisdom to inform a solid legal and policy framework
- Continue to engage Deaf community and build knowledge through funding→ i.e., Civic Innovation Grant

Conclusion

- AIxAI may never be able to replace humans due to interactivity issues (misunderstandings, etc)
- Aim for human-in-the-loop design

Authors

Teresa Blankmeyer Burke, Director, Gallaudet University School of Arts and Humanities and Professor of Philosophy and Bioethics

Jeff Shaul, Developer, GoSign.AI LLC

Holly Jackson, National Alliance of Black Interpreters, Atlanta

David Kerr, Executive Director, Canadian Association of Sign Language Interpreters

AnnMarie Killian, Chief Executive Officer, Telecommunications for the Deaf and Hard-of-Hearing, Inc.

Stephanie Jo Kent, Learning Lab for Resiliency®, ASL-English Community Interpreter

Guidelines Committee Recommendations

Interpreting SAFE AI Task Force Guidance (Ethical Principles): AI and Interpreting Services

- The Guidance is intended as a set of principles shared by all stakeholders, regardless of the setting or language
- The Task Force encourages stakeholders to develop setting- and country-specific Standards or Recommendations based on this Guidance.
  - E.g., Standards for AI Deployment in legal settings in the U.S., etc.
- The Task Force will review the Guidance at least annually or as needed.

SAFEAI

5 Ethical Principles

Principle 1: Adoption Prioritizes Accountability to End-Users

Accountable of AI technology for interpreting ensures that AI tools are procured andadoption utilized for interpreting services with explicit, opt-in informed consent, complete transparency, and adherence to ethical standards.

Principle 2. Improving Safety and Wellbeing

Creation of AI solutions for interpreting and incorporation of interpreting AI solutions in human communication must follow the existing legal and ethical frameworks for provision of interpreting services that are relevant for a particular setting of human communication or jurisdiction. If an AI solution is limited in its ability to meet standards of human interpreting, this limitation should be addressed either by not deploying this AI solution or making all parties aware of the limitations prior to the decision of utilizing it.

SAFEAI

5 Ethical Principles

Principle 3. Transparency of Technological and Interpreting Quality and Implementation

The principle of transparency refers to:
having policies and procedures that address the implications of using AI for interpreting as well the development of the AI-related tools for interpreting, communicating these and their implications to end-users.

The use of AI for interpreting should be disclosed to all parties. A disclaimer should be added as to what the key implications of using AI in the corresponding setting are.

Levels of Transparency:

- **For organizational purchasers of AI solutions**
- **For end-users of interpreting services**

5 Ethical Principles

Principle 4. Accountability

AI solutions should undergo validation by qualified human interpreters to establish a confidence level of accuracy prior to deployment. Liability for risks and harm associated with the use of AI solutions rests with the AI solution developers/vendors and organizations purchasing and deploying such solutions. Purchasers of AI solutions must establish quality assurance policies and procedures that explicitly define limitations of use and liability for misuse or non-disclosure of limitations.

Principle 5. AI as Part of Existing Interpreting Ecosystem

AI solutions for interpreting should follow ethical principles applicable to and expected of human interpreters in the field these solutions are deployed. We recommend that interpreters apply the same ethical considerations when they witness AI interpreting tools not meeting the ethical standards for their setting.

SAFEAI

Oct. 30
2023

OCTOBER 30, 2023

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. Purpose. Artificial intelligence (AI) holds extraordinary potential for both promise and peril. Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security. Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society.

Section 1557 Rule (45 CFR 92.201)

Updates to government regulations such as Section 1557 of the PPACA directly address AI and MT

§ 92.201 Meaningful access for individuals with limited English proficiency.

(c) Specific requirements for interpreter and translation services.

(3) If a covered entity uses machine translation when the underlying text 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, the translation must be reviewed by a qualified human translator.

Published May 6, 2024

ASTM F2575-23

ASTM F2575-23 has standardized the labels **BRT (bilingually reviewed translation) and UMT (unedited machine translation)** to provide LSCs and translation publishers with a means to be transparent by identifying the origin of their output and afford translation consumers some level of risk management.

> 10.2 The requester should consider labeling translations as either “Unedited Machine Translation” (UMT) or “Bilingually Reviewed Translation” (BRT) as information for the end users (see Annex A1). The requester then puts the translation(s) into use. If the end users are external to the requester organization, feedback may not be collected before the project ends, in which case it will be used for future projects.

Published January 1, 2023

Interpreting SAFE-AI Task Force

Bookmark
https://safeaitf.org/

Shape future of how interpreting services are delivered

Q&A and Closing Remarks