SAFEAI

SAFE-AI Task Force

October 2024

Interpreting with Integrity: A Deep Dive into Ethical AI

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.**

Emerging AI Legislation and Ethical Considerations

- **Data Privacy:** Ensuring that AI systems handle sensitive data securely.
- **Bias Control:** Addressing and mitigating biases in AI outputs.
- **Risk Consideration:** Evaluating the risks of AI use in high-stakes environments like DHS.
- **Language Pair Limitations:** Understanding which language pairs AI handles well and where human oversight is still necessary.
- **Transparency and Explainability:** Ensuring AI decision-making processes are understandable and can be audited, crucial for maintaining trust and accountability.
- **Human Oversight and Accountability:** Establishing clear protocols for human supervision of AI systems and defining responsibility for AI-assisted decisions and actions.

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

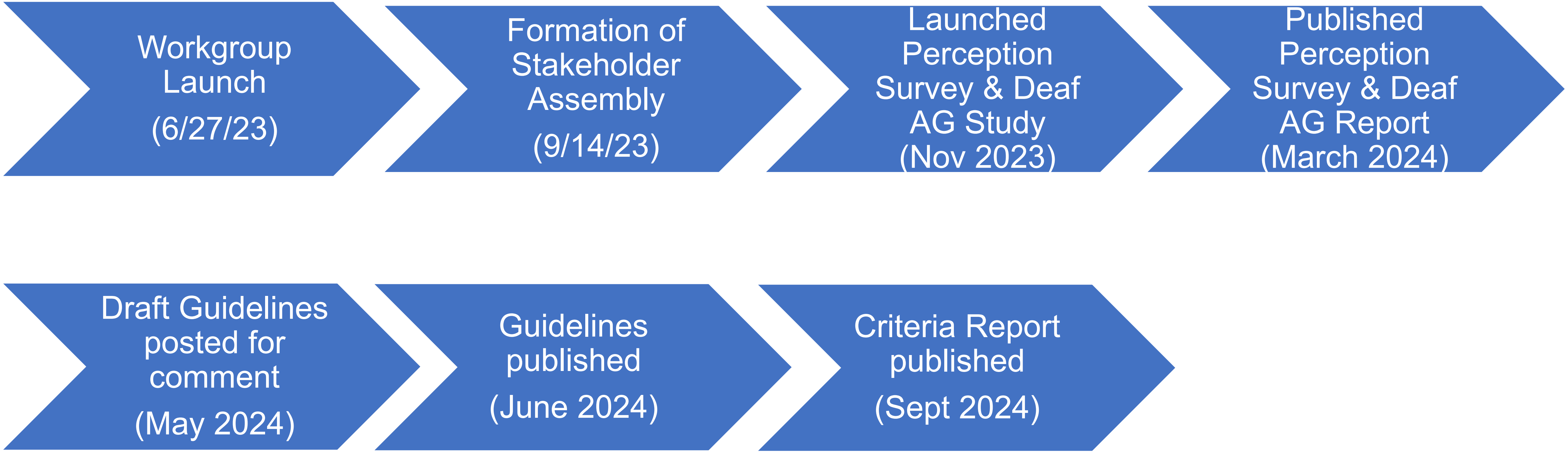
Dr. Bill Rivers, Principal, WP Rivers & Associates

SAFEAI

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

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)

Guidelines published

(June 2024)

Criteria Report published

(Sept 2024)

Perception Survey by the Numbers

- Conducted in 10 languages
- 2 Surveys (end-users and providers)
- 118 datapoints collected
- 2543 respondents
- 82 countries
- 48 states
- 79% based on the United States
- 3400 free text comments
- 9400 datapoints published

SAFEAI

Key Points: Perception Survey

- **90%+** of respondents do **not** trust interpreting from Apps provided by the organization or self procured Apps.
- **73%** of respondents fully trust face-to-face interpreters over other modalities of human interpreting.
- **9%** of respondents believe that automated interpreting can handle **simple conversations** and 25% believe that it can handle them soon.
- **1%** of respondents believe that automated interpreting can handle **complex conversations** and 8% believe it can handle is soon.
- Only **8%** agree that having machine interpreter is better than having no interpreter.
- Only **9%** stated that if they had to pay for an interpreter themselves, they would prefer a machine interpreter.

SAFEAI

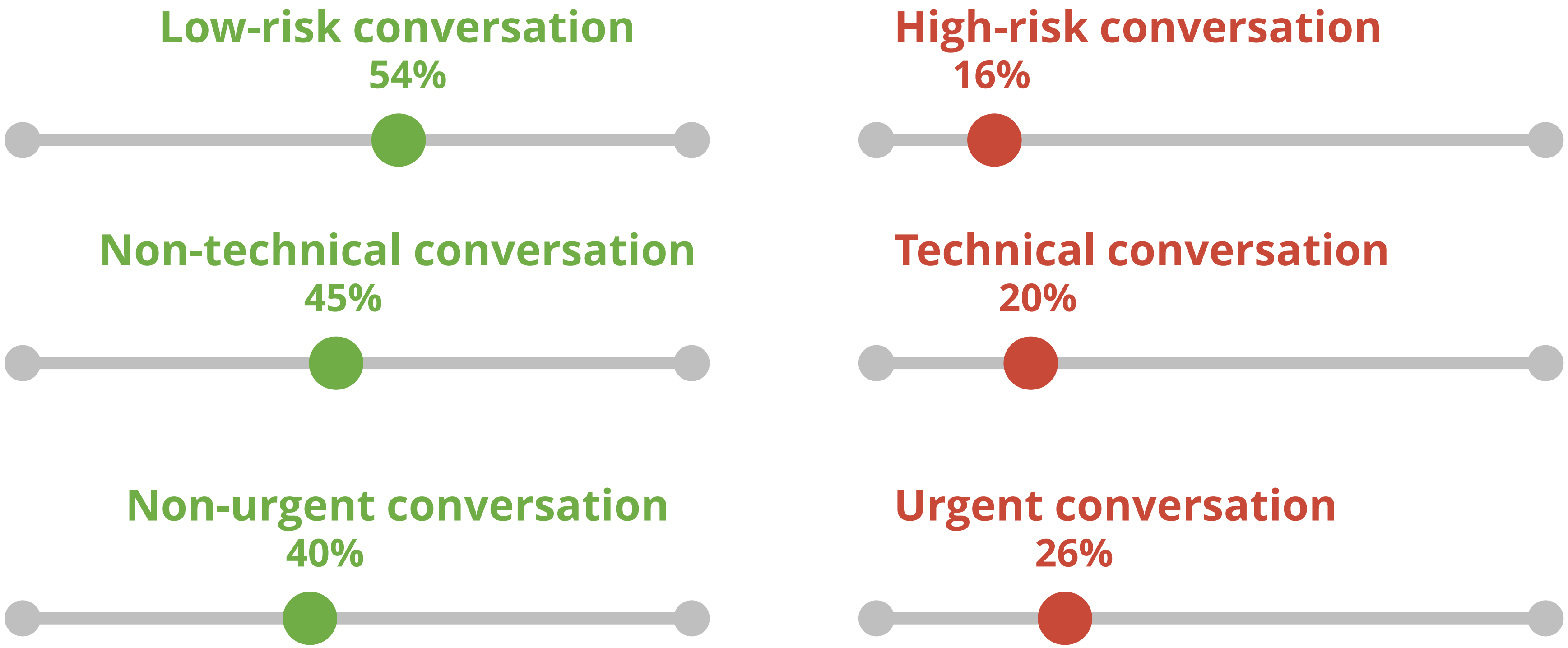
What Trips Up AI

- **Cultural Context:** cultural inferences and differences
- **Linguistic Context:** dialect, regionalism, colloquialism, slang, micro expressions, mixed languages, subculture, idiomatic expressions etc.
- **Interaction Context:** politics of the meeting, visual cues, intent, implicit information, etc.
- **Emotional Context:** attitude, tone of voice, mood, demeanor, feelings
- **Tone Context:** teasing, sarcasm, joking, innuendo, cynicism, humor, etc.

Imperfect Scenarios That Require Adaptation

- **Language Imperfections:** poorly expressed thoughts, incorrect grammar, half-expressed thoughts, mispronunciations, poor choice of vocabulary
- **Trauma:** victims of crime or abuse
- **Physical Impairments:** multi-disabilities, limb or hand difference
- **Cognitive Impairments:** older patient, active psychotic episode, low functioning
- **Speech Challenges:** speech impediments, strong accent, unusual cadence, strong accent, stroke patient, children's talk
- **Situational Barriers:** noisy background, multi speaker/signer, unusual positions, etc.

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



Low-risk conversation
54%

High-risk conversation
16%

Non-technical conversation
45%

Technical conversation
20%

Non-urgent conversation
40%

Urgent conversation
26%

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

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

Top 5 use cases
according to requestors and providers

#1 Notification (60%)
#2 Scheduling (57%)
#3 Balance inquiry (45%)
#4 Communication about a student absence (40%)
#5 Logistics handling (37%)
#5 Notification of a weather emergency (tie at 37%)

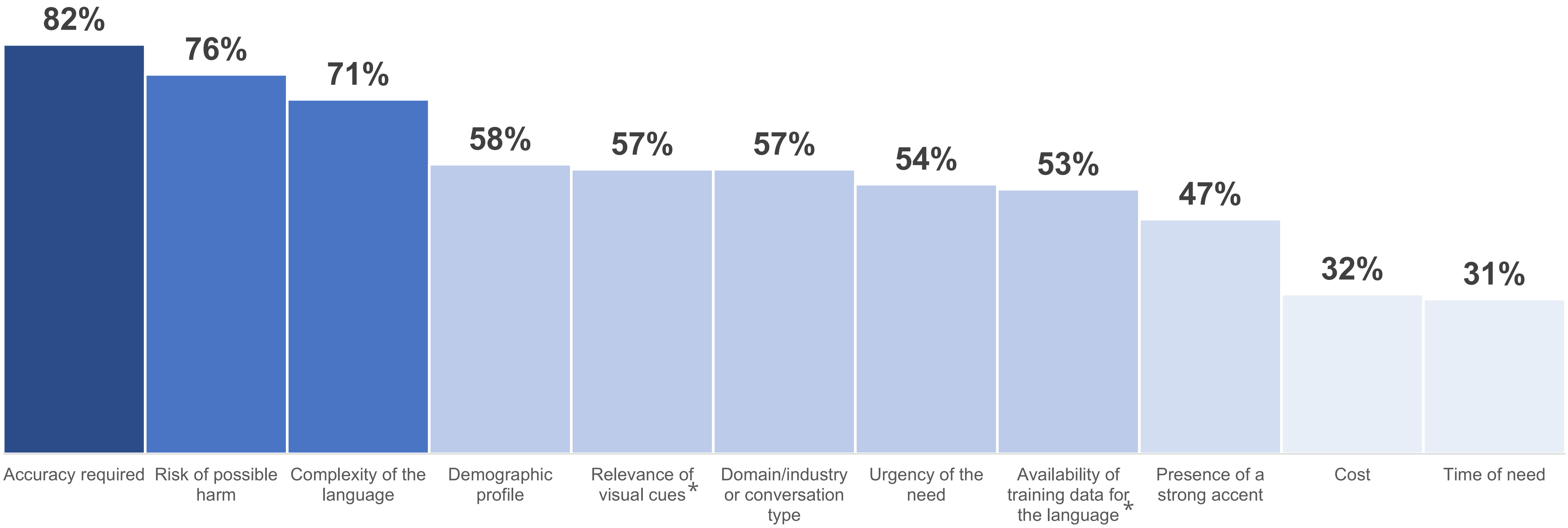
- **11 areas**
- **58 use cases**
- **~100 datapoints available in the report for each use case**

Percentage who think automated interpreting is mostly or totally suitable
Based on 504 to 1,819 requestors and providers, depending on the use case

SAFEAI

Based on “Perceptions on Automated Interpreting” © CSA Research

Which of the following elements are major criteria when deciding whether to use automated interpreting?



Based on 1,754 to 2,000 answers

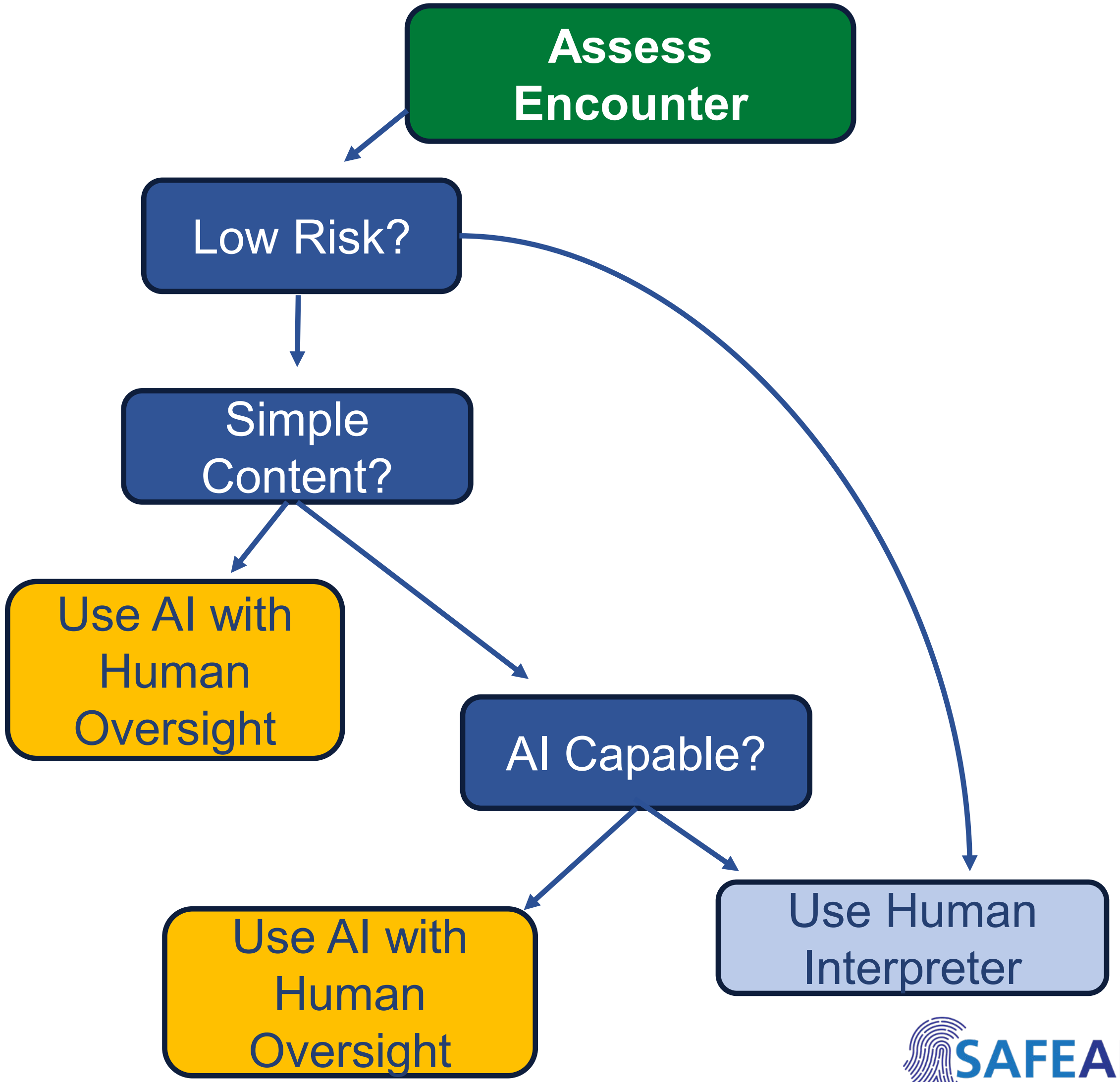
* *= End-users did not see this option*

SAFEAI

Based on "Perceptions on Automated Interpreting" © CSA Research

Conclusion: Decision-making Models are Essential

1. Start by **assessing the encounter**. Determine if the situation is **low risk**. If it's not low risk, use a human interpreter.
2. For **low-risk situations**, evaluate if the content is **simple**. If it's both low risk and simple, use AI interpreting.
3. For **low-risk but complex content**, **assess if AI is capable** of handling it. If AI is capable, use AI with human oversight.
4. If **AI is not capable**, use a human interpreter.



SAFE-AI Guidance: 5 Ethical Principles

1. **Accountability to end-users**
2. **Improvement of safety and well being**
3. **Transparency of technological and interpreting quality and implementation**
4. **Accountability**
5. **AI as part of existing interpreting ecosystems**

- Guidance available at www.safeaitf.org

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.

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.

Criteria Report: Automated Speech-to-Speech Interpreting (prepared by CSA Research – publication date September 2024)

Report examines situations in which organizations may use automated interpreting services to deliver language access

Evaluation dimensions:

- Benefit expectations
- Planned and unintended impacts
- Analysis of alternatives.
- Inputs into the session.
- Output expectations.
- Technology evaluation.
    - ➢ *Note: Technology evolves quickly. The recommendations in this report will change as technology matures and end-user perceptions evolve.
    - ➢ ** Note: Not all languages are supported by technology.

Report concludes with 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.

How AI can help if it’s not ready to interpret

Computer-aided interpreting (CAI)

- Terminology research
- Automated term lookup
- Note-taking
- Understanding accents

Validation of human work

- Interpreting training
- Cross-checking of human accuracy (e.g. important elements in trial to make sure they were interpreted correctly)

AI validated by human

- Live
- After the fact

Q&A