Vendor Evaluation Guide: AI Captioning and Translation for Meetings with Deafblind Participants

As AI continues to evolve, you may have noticed rapid growth in automated captioning and a growing curiosity about real-time AI translation during meetings. For a global community, these innovations are exciting and full of possibility, but they also come with important considerations. This resource has been created to help teams navigate those conversations with the needs of the deafblind community at the centre.

The document is divided into both technical and non-technical sections so that anyone regardless of technical background, can participate meaningfully in decisions about using or purchasing these tools. It isn't meant to be a definitive framework or a checklist of requirements. Instead, it's a practical guide to support thoughtful discussion, informed investment, and shared understanding as organizations explore AI-enabled accessibility solutions.

This work is released under a Creative Commons Licence to encourage adaptation, collaboration, and collective learning. Your expertise and lived experience are essential as we continue to refine best practices and ensure these technologies genuinely enhance access, inclusion, and participation for the deafblind community.

Want to contribute or ask a question? Drop us a line at dbitechnetwork@gmail.com.

Contents

HOW TO APPLY THIS GUIDE THROUGHOUT THE EVALUATION PROCESS

Most vendor demonstrations are brief, often around 30 minutes, which is rarely enough time to explore everything needed for a thorough, accessibility-centered evaluation. This resource is intentionally comprehensive because it reflects the complexity of AI-enabled accessibility tools. You are not expected to cover every item during a short meeting. Instead, the demo should be viewed as one component of a broader, thoughtful evaluation process.

To ensure the demo supports an informed, accessibility-centered evaluation, it helps to begin with a simple structure. The steps below outline a balanced approach that respects limited meeting time will still honouring the needs and perspectives of Deafblind users.

1. Define the use case, and business requirements
2. Conduct preliminary research
3. Research the vendor's credibility and stability, and identify a set of priority questions
4. Use the meeting to assess clarity and transparency
5. Include lived experience
6. Complete the post analysis

1- Define the use case and business requirements

A clear understanding of the use case and business requirements provides the foundation for a meaningful evaluation. Prior to engaging with vendors, teams should take a moment to articulate what they are trying to achieve, who will be using the tool, and what conditions must be met for the solution to be viable.

Use case or objective

Identify the specific problem the tool is intended to solve and the real-world scenario in which it will be used. Clarify the primary users, including Deafblind users, and the tasks or workflows the tool needs to support.

Business requirements

Outline the practical constraints and expectations that shape the evaluation. These may include budget, licensing model, security and privacy needs, data handling expectations, integration with existing systems, and required output formats.

Accessibility and inclusion requirements

Document the non-negotiable accessibility expectations from the outset. This includes compatibility with assistive technologies, support for Deafblind users, clarity of documentation, and expectations for transparency around limitations and known issues.

Success criteria

Define what a successful outcome looks like. Identify the must-have requirements, the nice-to-have features, and the indicators that will help you determine whether the tool meets the requirements.

2 - Conduct preliminary research

Before meeting with a vendor, a small amount of preparation can save significant time and ensure the conversation focuses on what truly matters. A quick review of publicly available information helps you understand the basics, reducing the need to spend meeting time on details that can be gathered in advance.

What to review

Begin by reviewing the vendor's public website or 'Trust Center,' as many compliance documents are readily available for download. If any items are missing, request the most recent versions directly from your account representative to ensure you are evaluating current data and evaluate the company's legitimacy.

Tip: Searching for the company name + 'VPAT' or 'Trust Center' on Google is often faster than navigating their website menus.

- **Privacy policy and data processing agreement**
  Look for clear statements about data ownership, retention, and whether users can opt out of contributing to the model training.
- **Accessibility documentation and WCAG 2.2 compliance**
  Confirm that accessibility audits are conducted by qualified experts, including members of the disability community, and that the product aligns with current Web Accessibility standards and guidelines: https://www.w3.org/WAI/standards-guidelines/wcag/
- **Supported assistive technologies**
  Ensure the vendor explicitly confirms compatibility with NVDA, JAWS, VoiceOver, and refreshable Braille displays.
- **Accuracy and performance information**
  Request sample accuracy reports, including performance in noisy environments and with diverse accents. Even high-level accuracy claims (e.g., 95%) can be

misleading as an error of 1 out of every 20 words is enough to alter meaning or context.

- **Product Roadmap**
  Review plans for the next 12 – 24 months, especially commitments related to accessibility improvements. Given how quickly AI evolves, a tool that is "mostly accessible" today should demonstrate a clear path toward becoming fully accessible in the future.
- **Invite a Deafblind community member** to participate in the meeting or review materials in advance so the evaluation includes direct lived experience and meaningful feedback.

3- Research the vendor's credibility and stability, and identify a set of priority questions

In addition to reviewing product-specific documentation, it is helpful to understand the vendor's overall maturity and reliability.

- **Company history and stability**
  How long has the vendor been in business, and what is their track record of maintaining and improving their products?
- **Product maturity**
  Is the tool newly launched, in beta, or well-established in the market?
- **Customer base**
  Are there examples of organizations currently using their product, particularly those with similar accessibility or compliance requirements?
- **Testimonials or references**
  Are there public case studies, customer stories, or an opportunity to speak directly with an existing client?
- **Shortlist 2–3 vendors that:** advertise braille/display compatibility or hybrid human captioning, offer trials, and provide clear Data Processing Agreement.

Identify a set of priority questions

Identify three to five questions that reflect your must-have requirements and the specific needs of Deafblind users. This keeps the discussion focused and ensures the most important topics are addressed first.

4- Use the meeting to access clarity and transparency

In a short demo, you may not see every feature, but you can observe how the vendor communicates. Do they answer questions directly? Do they acknowledge limitations?

Do they demonstrate an understanding of accessibility and lived experience? These cues often reveal as much as the technology itself.

5- Include lived experience

Ask about short trial access

Requesting a short trial access allows you to assess the tool in real-world conditions rather than relying solely on vendor descriptions or marketing materials. A trial gives you the opportunity to test the AI captioning and translation features with your actual content, workflows, and accessibility needs.

It also ensures that a person with lived experience can meaningfully evaluate usability, accuracy, and accessibility barriers that may not be visible through documentation alone. Their feedback helps us identify strengths, limitations, and potential risks early in the process.

Request a meeting follow-up

The follow-up meeting with the vendor is an intentional and necessary step in your evaluation process. While this document supports your inquiry into AI captioning and translation tools, several key criteria, such as accessibility compliance, accuracy expectations, and real-world usability, are most effectively examined through direct discussion. This is particularly important once a person with lived experience has tested the trial version and provided their feedback.

It is also important to recognize that taking the time to ask detailed questions, seek clarification, and verify vendor claims is a standard and expected part of a responsible, accessibility-focused evaluation process. Vendors should be prepared to support this level of due diligence.

Furthermore, meeting with the vendor enables you to validate the information they have provided, address any gaps or ambiguities, and ensure there is a shared understanding of their capabilities. This step supports a well-informed, evidence-based decision and reduces the risk of misalignment during procurement or implementation.

Other tips: What to bring to the meeting (if available)

- A custom glossary of technical terms
- Sample audio with different accents
- List of your required integrations (e.g. learning management tools, webinar platform)

6- Conduct post analysis

Follow-up emails and written responses

Requesting written follow-up from vendors is an important part of your evaluation process. Written answers allow you to conduct a thorough post-meeting analysis, compare vendors consistently, and share information with colleagues who were unable to attend. This also gives vendors the time and space to provide accurate, thoughtful, and complete responses rather than reacting under pressure during a live call.

Key Considerations

Explore more compatible options if:

- There are no options to schedule a demo live with assistive technology
- Vague answers about where data is stored
- No testing with users with lived experience
- Claims "100% accuracy"
- Cannot provide WCAG compliance
- No opt-out from AI training data
- The demo only works with prepared samples
- Captions lag more than 5 seconds
- Cross talk in multiple languages. It is the vendor responsibility to demonstrate how their system performs under these conditions
- Can sessions be time-limited, and how is billing handled when a user is idle? For example, if a meeting ends after one hour but the session remains open for several hours without activity, is billing based on active use or total elapsed time? (if based on active use, can ambient noise trigger billing?)

It is essential to remember that the demo is only one step. The deeper assessment happens through hands-on testing, documentation review, and team discussion, especially once individuals with lived experience have had the opportunity to evaluate the tool and share their insights. There is no expectation to make a decision during or immediately after the meeting.

TECHNICAL SECTION

Technical Testing / Questions/ Demo

Invite your team to review the guide together and identify which items are essential, recommended, or optional, using the Vendor Scorecard to support consistent evaluation.

1. Caption Accuracy

a) Test with different accents utilizing technical vocabulary and content, as well as casual conversation

b) Check speaker labeling (every new line of text should be preceded by the speaker's name (e.g., "Sarah: [Text]"); When the speaker changes, does the assistive technology announce the new name, or does the user have to scroll back to find it?; Does the labeling include important context like **[Laughter]** or **[Cross-talk]** so the user knows why there might be a pause in the conversation?)

c) Measure lag time (should be under 5 seconds)

d) Ask: How does accuracy change with diverse accents?" (Many AI models are often trained on "Standard American English." If a speaker has a regional accent, speaks quickly, or uses a different dialect, the accuracy may drop.)

e) Ask: "Which terms did the system struggle with?" or "Does the system provide a 'Confidence Score'?" (some advanced tools can flag words they are unsure about, letting the user know the translation might be "fuzzy" in that moment)

2. Deafblind User Experience

a) Test with screen reader (JAWS, NVDA, VoiceOver)

b) Test braille display output

c) Test keyboard-only navigation

d) Test caption customization (font, contrast, placement)

e) Ask: "How are speaker changes announced to screen readers?"

f) Ask: "Can captions route directly to braille displays?"

3. Translation Quality

a) Test with 3+ languages simultaneously

b) Check translation lag (should be under 5 seconds)

c) Ask: "How does the technology handle terms without direct translations?"

4. Minimum System Requirements

a) Test with slow or unreliable internet (high latency)
b) Test with background noise
c) Test with multiple simultaneous speakers
d) Ask: "How does the product respond to interruptions in connection?"
e) Ask: "What is the minimum bandwidth required?"
f) What are the minimum system (computer) requirements

5. Privacy, Data, and Security

a) Check retention and deletion settings
b) Ask: "Where is data stored (which countries)?"
c) Ask: "Is our data used to train your AI? Can we opt out?"
d) Ask: "Who has access to our meeting data?"
e) Ask: "What happens to our data if we cancel?"
f) Ask: "Do you share data with any external or 3rd party services
g) Ask: Which privacy regulations do you comply with? Eg: GDPR, PIPEDA, CCPA? Contractually require opt-out or data deletion.
h) Ask: Does the organization align its security practices with any established standards (SOC, ISO 27001, NIST, CIS, or equivalent)
i) Ask: Do you undergo regular third-party security audits or penetration tests

6. Integration and workflow

a) Install and activate the product (count the steps)
b) Ask: Can I generate a 2-minute transcript (TXT file) during the live demo. Open the TXT file immediately. Check: does the export look like a normal document, or is it cluttered with numbers and symbols? If it's hard to read, the tool isn't truly accessible for post-meeting review.
c) Check: In the exported file, is each line of text preceded by the speaker's name? Ask: "Can I see a sample of how a speaker change looks in the code?"
d) Ask: "Does your platform have a built-in editor where I can fix mistakes *before* downloading the SRT/VTT file? If I fix a word in the transcript, does it automatically update the timestamps in the subtitle file?"
e) Test integration with your Learning Management System if applicable
   o Many integrations put captions inside an "iframe" (a window within a window). Screen readers often struggle to "jump" into these windows, making the captions invisible to assistive tech.

- If enabling captions in the LMS requires using a mouse, a deafblind user relying on keyboard navigation or a Braille display may be unable to activate the captioning feature.
- If the LMS or the captioning and translation product experiences delays, the captions may fall out of sync with the video, making it difficult to follow the lesson in real time.

f) Ask: "When the file moves to your system, does it keep the labels (e.g., "Speaker name: [Text]") or does it flatten the text into one long, confusing paragraph?

g) Ask: "Do breakout rooms allow translation continuity?"

After the Demo

Frequently asked follow-up questions:

This section outlines common follow-up questions that organizations often raise after the initial meeting. It provides a helpful starting point as you prepare your own inquiries. You may select questions from the earlier section or use the ones listed below, depending on what you were able to cover during the initial conversation and what remains relevant to your organization's needs.

- Data and privacy processing agreement for review
- Pricing breakdown and payments.
- Cancellation and Refunds
- Renewals (are they automatic?)
- Customer references (similar organizations)
- 30-day trial with your team
- If applicable, schedule another meeting to test their software with a person with lived experience
- Service Level Agreement (SLA)
  - Defines uptime, and availability of the product
  - Defines vendor support levels, methods, and times (E.g. Mon-Fri 8AM – 6PM, and 24/7 for critical issues etc.)
  - Defines notification requirements regarding planned outages and scheduled maintenance
  - Defines performance expectations and metrics
  - Defines penalties if the vendor does not meet or exceed the SLA
  - Clarity on SLA availability metrics. E.g.: monthly, annually?

Score the Vendor (0-5 each)

While this scoring table provides a helpful foundation, it isn't exhaustive. You are encouraged to tailor these questions to align with your organization's requirements and user priorities. The scoring is inherently subjective and should be guided by what matters most to your teams and communities.

Use the 0–5 scale to assess how well each vendor aligns with your needs, recognizing that the purpose of this tool is to support thoughtful, context-aware decision-making rather than to produce a definitive or universal rating. A suggested scoring scale is available below to assist with rating each criterion.

Vendor Accessibility and Performance Scorecard

Date: **Minimum to proceed:** 28/35

Vendor info:	
Reviewed by:	

Area	Score	Reasons for this score
Caption accuracy on your content		
Braille display compatibility and other Assistive Technology Works with NVDA/ JAWS/ Voiceover		Test: braille routing, screen reader flow, speaker labels, noisy room, multi-language.
Translation quality/ Caption lag <5s		
Data residency & encryption		
Integration ease (Zoom, LMS, etc)		
Vendor knowledge of accessibility		
Upload custom glossary		
Option for CART/human scribe		
Other		
TOTAL	**/40**	

0–5 Scoring Scale

Score	Meaning	Description
0 – Does Not Meet Requirements	Fails to address the criterion	The vendor provides no evidence, the feature is missing, or the response contradicts your needs.
1 – Below Expectations	Major gaps	The vendor offers minimal or unclear information; substantial concerns remain.
2 – Partially Meets Requirements	Some elements present	The vendor addresses the criterion inconsistently or with notable limitations. Additional clarification or testing is required.
3 – Meets Requirements	Acceptable performance	The vendor provides sufficient evidence and functionality to meet baseline expectations. Some minor gaps may remain.
4 – Exceeds Requirements	Strong performance	The vendor demonstrates clear strengths, strong evidence, and reliable functionality with few concerns.
5 – Fully Meets and Surpasses Requirements	Exceptional performance	The vendor provides comprehensive, high-quality evidence and demonstrates outstanding alignment with your needs, including accessibility, usability, and sustainability considerations.

Essential accessibility criteria

Consider:

- WCAG 2.2 Level AA compliance (AAA preferred)
- Works with 2+ screen readers
- Data encryption (transit and at rest)
- Clear data deletion policies
- 95%+ accuracy on your technical terms
- Translation latency under 5 seconds
- 99.5%+ uptime guarantee

- - Some vendors exclude "planned maintenance" from their uptime math. If the system is down for 5 hours every weekend, their real-world reliability is much lower than 99.5%.
  - Ask for their Public Status Page; this site should show a 90-day history of every crash and delay.
- Positive reference from similar organization
- Plan for hybrid coverage. AI for routine meetings; CART or human Scribes for high-stakes sessions and when tactile (braille) fidelity is essential.

How to use the scoring scale in team discussions

Using the 0–5 scale during team discussions helps ensure that evaluations are consistent, transparent, and grounded in shared criteria rather than individual impressions. The goal is not to force agreement but to create a structured way to surface insights, identify gaps, and understand where perspectives differ.

1. Begin with individual scoring

Each team member should score the vendor independently before the group discussion. This prevents early opinions from influencing others and ensures that lived-experience feedback is fully represented.

2. Compare scores to identify alignment and divergence

Once individual scores are shared, look for:

- Areas where scores cluster closely (indicating shared confidence)
- Areas where scores vary widely (indicating uncertainty, different experiences, or missing information)

Both patterns are valuable. Divergence often reveals questions that need clarification or features that require deeper testing.

3. Discuss the reasoning behind the scores

Encourage team members to explain:

- What evidence informed their score
- What worked well or poorly during testing
- Any accessibility or usability barriers they observed
- Any assumptions they made due to missing information

This step ensures that the conversation honours lived experience and brings multiple perspectives into the decision-making process.

4. Identify follow-up questions or testing needs

If the discussion reveals gaps, unclear vendor claims, or inconsistent performance, note these as:

- Questions for the vendor
- Areas requiring additional testing
- Criteria that need more evidence before scoring is finalized
- This keeps the process iterative rather than rushed.

5. Use the scores as guidance, not a final verdict

The scale is a tool to support decision-making, not a rigid formula. A lower score may highlight a risk that needs mitigation; a higher score may indicate a strength worth prioritizing. The final decision should reflect:

- Organizational priorities
- Lived-experience feedback
- Technical feasibility
- Long-term sustainability

You can also:

6. Document the rationale

Record both the scores and the reasoning behind them. This ensures transparency, supports future reference, and allows colleagues who were not present to understand the decision-making process

NON-TECHNICAL SECTION

Non-technical vendor comparison table

The following is a non-technical vendor comparison table designed to help teams evaluate AI translation and captioning tools side by side. To start, refer to "How to apply this guide through the evaluation process"

Invite your team to review the guide together and identify which items are essential, recommended, or optional, using the Vendor Scorecard to support consistent evaluation.

Category	Evaluation Criteria	Vendor A	Vendor B	Vendor C
Accessibility	Interface is simple and easy to navigate, including keyboard shortcuts			
	Buttons and controls are clearly labeled (every interactive button or menu such as play, pause, settings, or language selection has a meaningful, descriptive label that can be read by screen readers and displayed accurately on braille devices.)			
	Caption customization (size, color, font, placement)			
	Works with screen readers and other assistive technology			
	Tested with people with lived experience, e.g. deafblind, low vision			
	Follows the WCAG 2 Overview | Web Accessibility Initiative (WAI) | W3C			

Category	Evaluation Criteria	Vendor A	Vendor B	Vendor C
Technical	Works reliably on standard internet bandwidth			
	Works with regular meeting room microphones			
	Handles background noise and group discussions well			
	Compatible with Zoom, Teams, hybrid setups			
	Supports breakout rooms			
Language & Accuracy	Handles accents and dialects effectively			
	Caption/translation delay (5 seconds approximately lag time)			
	Accuracy with specialized terminology			
	Ability to upload custom glossaries			
	Strategy for untranslatable terms (some words, phrases, or cultural concepts don't have a direct equivalent in another language, and the vendor should have a clear plan for how their system handles these events)			
	Transparency about expected error rates			

Category	Evaluation Criteria	Vendor A	Vendor B	Vendor C
Privacy & Security	Where data is processed and stored. What to consider: 1- Countries have different privacy laws 2- Some organizations must keep "data residency" within the country where the product is utilized Storage could include: • Audio recordings • Captions or transcripts • Translation output • Metadata (timestamps, meeting ID, device type)			
	Option to opt out of AI training			
	Data retention policy is clear (what happens to your audio, video, captions, and translations after the meeting ends)			
	Access controls are documented (the vendor should clearly explain who can access your data, under what conditions, and how that access is restricted)			

Category	Evaluation Criteria	Vendor A	Vendor B	Vendor C
Privacy and Security	Compliance with privacy standards (common ones include: • GDPR (European Union) – protects EU personal data and strengthens user privacy rights. Applies to any entity worldwide processing data of EU residents. • PIPEDA (Canada) – governs how private-sector organizations handle personal information. • HIPAA (United States) – protects health information. • CCPA/CPRA (California) – grants consumers rights over their personal data. • FERPA (United States) – protects student information. • Local or sector-specific rules (public sector, education, healthcare, Indigenous data sovereignty.)			
	Ability to delete all data on request, or/ and contractually require opt-out or data deletion.			
Ethical Considerations	Approach to reducing gender/cultural bias			
	Handling of domain-specific language Examples include: • Medical: "triage," "contraindicated," "informed consent" • Legal: "affidavit," "liability waiver," "due diligence" • Technical: "API," "bandwidth," "encryption key" • Education: "IEP," "credit recovery," "accommodations" • Community services: program names, local terminology, cultural concepts			

Category	Evaluation Criteria	Vendor A	Vendor B	Vendor C
	Transparency about known limitations (the vendor explains where the tool may struggle, including language, cultural, technical, and environmental limitation. Users understand when accuracy might drop and what to expect in real-world use).			
Licensing & Usage	Pricing model (per minute, per event, per user)			
	Limits on simultaneous users (performance of the tool in small and large events)			
	Enterprise licensing options			
	Hidden or additional costs			
Performance & Reliability	Tool behavior during outages			
	Uptime Service Level Agreement or SLA (what happens if your tool goes down during a meeting? Some vendors may offer: • Service credits • Refunds • Priority support • Escalation pathways			
	Real-time monitoring or alerts (Admin is notified when: • captions or translation slow down • audio quality drops • the system is experiencing delays • servers are overloaded • the service is about to go down • there is a temporary outage			

Category	Evaluation Criteria	Vendor A	Vendor B	Vendor C
Support & Operations	Response time for critical issues			
	Live support during events			
	Onboarding/training provided			
	Availability of documentation and tutorials			

Author:

Carolina Cohoon, Leader Inclusive Learning Design. Co-Chair DBI Technology Network

Acknowledgements:

The creation of this document has been informed by collaborative input. I extend my appreciation to those who provided insights and critical review:

Chris Cohoon, Senior IT Leader. Digital Transformation

Jeremy Nathan Marks, Researcher

Mia Sellars, Candidate, Master of Occupational Therapy

Dr. Nicole Lanthier, Audiologist

Michael Pyett, Senior Technology Leader

Roger Wilson- Hinds, Trustee Deafblind UK (Retired Head Teacher)

References:

https://www.w3.org/WAI/standards-guidelines/wcag/

https://e-channel.ca/practitioners/resources/ai-bytes/

Professional & AI-Based Captions for Deaf & HoH | Ava