nagarro

Is AI bias toward disabled people an accessibility issue?

Fluidic Intelligence

Contents

Executive summary

AI increasingly determines who gets hired, how people access healthcare, how students are evaluated, and how individuals interact with essential digital services. But when these systems are trained on incomplete, skewed, or ableist data, they misinterpret disability-related patterns, misread assistive technology, or erase disability entirely. These issues are not simply model errors – they are accessibility failures that silently block disabled people from equitable participation.

As decision logic moves deeper into algorithms and away from visible interfaces, accessibility must evolve from focusing purely on *usable interfaces* to ensuring *fair outcomes*. Equitable AI is no longer optional; it is a foundational requirement for responsible, safe, and inclusive technology. This paper demonstrates why equitable AI outcomes must become a core accessibility mandate and offers concrete steps to begin that transformation.

What you'll gain from this paper

A comprehensive reframing of accessibility that includes AI-driven decision-making

A detailed model of how disability bias emerges across data collection, labeling, model design, and deployment

Practical methods – metamorphic, adversarial, error-guessing, and explainability testing – to uncover hidden barriers

Guidance on building representative datasets, accessible appeal mechanisms, and human-in-the-loop safeguards

A blueprint for embedding disability inclusion into every stage of the AI lifecycle

Key findings from real-world analysis

- LLMs consistently associate disability terms with negative sentiment, reinforcing harmful narratives
- Major speech recognition tools show accuracy gaps of 30%+ for atypical or disfluent speech
- Hiring and biometric systems misclassify disability-related behavior as "risk," "fraud," or "poor fit"
- Image-based AI tools erase mobility aids and visible differences, normalizing able-bodied defaults

Introduction

AI isn't just in our phones, search engines, and smart speakers; it increasingly mediates who gets hired, how students are assessed, what care patients receive, and how we navigate digital services each day. Its decisions shape opportunities and outcomes. But what happens when those AI-based decisions, often trained on incomplete or biased data, misinterpret disability, misread assistive tools, or reinforce stereotypes? The result is not only unfairness but also an *accessibility barrier.* When an inaccessible interface blocks a user, we call it an accessibility violation. Similarly, when an AI outcome blocks people with disabilities from fair consideration or equitable service, *we should call that too as an accessibility failure.*

Expanding the scope of accessibility

Accessibility has traditionally focused on user interfaces: readable type, keyboard navigation, captions, color contrast, and compatibility with assistive technologies. That work remains essential. Yet as decision-making shifts from visible user interfaces to AI-based models, accessibility must expand its scope, from “Can I use it fairly?” to “Will it treat me fairly?” This reframing recognizes that inclusion is not merely about perceiving and operating content at UI; it is also about receiving equitable inclusive outcomes from AI models. Bias in AI, whether through training data, model architecture, labeling practices, or deployment context, can exclude disabled people in quieter but more consequential ways than a missing Alt Text tag ever could.

Some recurring patterns of AI bias towards disabled users

- **Misinterpreting the use of assistive technology as suspicious activity:** For example, when a person uses a screen reader or other accessibility tool, AI systems may wrongly flag their navigation patterns as “bot-like” or fraudulent because they differ from typical mouse or touch behavior.
- **Ignoring or erasing disability-related context in training data of AI model,** which makes AI models brittle when encountering atypical patterns (e.g., non-standard speech, movement, or interaction timing).
- **Reinforcing stereotypes through indirect signals:** For instance, AI systems may treat things like gaps in employment history or irregular work patterns as signs of poor performance. But, in reality, they could be due to medical treatment or disability-related leave.

When these patterns drive decisions about jobs, services, or safety, we cross the line from inconvenience to exclusion. That is an accessibility issue.

Understand the scale and nature of the problem

Scale of disability: Over **1.3 billion people**
(about **16% of the global population**) live with a disability (World Health Organization, 2022).

Voice and speech systems

1. AI Natural Language Processing models show disability bias

Researchers at Pennsylvania State University (2022) analyzed 13 major language models and found a consistent negative bias toward disability-related terms. Words like *"autism," "wheelchair,"* or *"mental illness"* were often linked with negative emotions or language, such as "suffering" or "burden." This demonstrates how AI text systems can unintentionally reflect ableist stereotypes, portraying disability as something undesirable rather than a normal part of human diversity Impact: AI systems propagate harmful stereotypes about disability, reinforcing their portrayal as problems rather than human diversity.

*Source: Penn State College of Information Sciences and Technology – **"AI language models show bias against people with disabilities."***

2. Speech recognition systems discriminate against speech disabilities

A 2024 study from Michigan State University evaluated six leading Automatic Speech Recognition (ASR) systems (including Google, IBM, and Microsoft) on speech from people who stutter. Every ASR tested exhibited statistically significant accuracy bias, with high word error rates (WER).

Example: Speech disfluency users experienced accuracy gaps of over 30% compared with "non-disabled" controls.

*Source: arXiv.org (2024) – **"Lost in Transcription: Identifying and Quantifying the Accuracy Biases of Automatic Speech Recognition Systems Against Disfluent Speech"***

Hiring algorithms

AI recruitment systems penalize disabled candidates

Multiple reports – including the University of Melbourne (2025) and Ford School of Public Policy (2024) – revealed that AI recruitment tools systematically downgraded disabled candidates. For example, software analyzing facial expressions, tone, and microgestures failed to account for motor or speech impairments These tools often misinterpreted or ignored atypical speech, facial movement, or slower response times caused by assistive devices.

*Sources: AI bias in hiring, AI Multiple (2025) – **"Bias in AI"**; Michigan Ford School of Public Policy (2024) – **"Artificial Intelligence (AI) Hiring Technology and Disability Discrimination"***

Biometric bias

Facial recognition systems flag disabled faces as "Unsafe."

Facial recognition systems and content filters have been documented to blur or remove images of people with visible facial differences, misclassifying them as "graphic" or "unsafe content."

Examples include individuals with cleft palate, craniofacial syndromes, or disfigurements being censored on social platforms.

Report: Interface UK (2023) and Sheri Byrne-Haber (registered accessibility researcher) showed AI content moderation frequently mislabels such users, worsening digital exclusion.

*Sources: Interface.org.uk – **"Artificial Intelligence and Facial Discrimination"**;*
*Sheri Byrne-Haber – **"Disability and AI Bias"***

AI "erasing" disability

First-person documentation from *Time Magazine* (2025) described a case where AI image editing tools automatically removed mobility aids like wheelchairs or braces in photos, reinforcing aesthetic ableism by normalizing invisibility of disability.

*Source: Time Magazine (June 2025) – **"When AI Erased My Disability"***

These facts point to a central reality - AI systems often generalize poorly for disabled users because disability is underrepresented, misrepresented, or treated as noise in datasets and labeling pipelines. When models encounter patterns, they never learned to value, they fail in ways that map directly to accessibility harm.

Why traditional accessibility isn't enough

Classic accessibility practice excels at ensuring perceivability, operability, understandability, and robustness at the user interface level. But if the underlying decision logic is biased, even a beautifully accessible interface can deliver discriminatory outcomes. Accessibility must therefore encompass both:

- **User Interface access:** Ensuring assistive technology compatibility, clear content, and inclusive interactions.
- **AI outcome access:** Ensuring that AI algorithmic decisions do not exclude or degrade service for disabled people.

This shift reframes accessibility as a socio-technical commitment. It asks: *Is the path through the product accessible, and are the decisions along that path fair?*

What causes AI bias for disabled users in AI systems/models/algorithms?

Data gaps and skew

Disability is often under-represented, stigmatized, or stripped out during data collection. The result: Models learn “normative” interaction patterns and penalize divergence.

Labeling bias

Annotators may misclassify atypical speech as “noisy,” atypical gaze as “deceptive,” or assistive input patterns as “bot-like.” These biased labels encode stereotypes into the model.

Proxy discrimination

Even when disability is not explicitly included, correlated features (employment gaps, test-taking time, input modalities, or error rates on speech recognition) become proxies that drive adverse decisions.

Context loss in deployment

Systems tuned in lab settings often fail in real-world contexts where disabled users’ variability is higher. Without monitoring and feedback loops, error rates remain hidden within aggregate metrics.

Feedback barriers

Disabled users may face higher friction in appealing an automated decision, especially when explanations are opaque or inaccessible.

Real-world data supporting these causes

Disability data missing from algorithmic tools

The Center for Democracy and Technology (CDT, 2024) reported that most commercial AI datasets lack disaggregated representation of disability, leading to biased outcomes in healthcare, welfare, and employment algorithms.

Example: AI decision systems assessing job eligibility or treatment access heavily penalize atypical movement or response patterns.

*Source: Forbes – **"Disability Data Alarmingly Absent From AI Algorithmic Tools"** (2024).*

Sociotechnical ableism in AI (Academic Lens)

A 2025 peer-reviewed article titled Disabling AI: Power, Exclusion, and Disability (Taylor & Francis) describes "sociotechnical ableism" – how exclusionary data, design, and institutional norms reproduce able-bodied assumptions in machine learning systems.

The concept relates to structural power in tech design, highlighting how datasets omit non-normative bodies or behaviors.

*Source: Foley, A. (2025). **"Disabling AI: Power, Exclusion, and Disability"***

What counts as an "Accessibility Failure" in AI?

Borrowing from traditional accessibility logic: if a barrier prevents a person with a disability from effectively accessing a service, it is an accessibility failure. In AI **we have some indicators as well.**

Discriminatory outcomes

When AI systems make unfair decisions that disproportionately reject, penalize, or misinterpret disabled users (e.g., job screeners rejecting applicants due to disability-related behavior or history).

Misrecognition of disability signals

When tools like speech or facial recognition fail to interpret atypical voices, movements, or expressions, they lead to frequent errors or exclusion from services.

Inaccessible automated processes

When users cannot understand, challenge, or appeal an AI-driven decision because explanations are too technical or not provided in accessible formats.

Lack of representative data

When disability-related patterns or assistive technology interactions are missing from training data, causing AI systems to "break" when encountering them in real life.

Barriers created by security or fraud detection systems

When assistive technology usage (like screen readers or eye trackers) is flagged as suspicious, blocking access to accounts or online forms.

Absence of human fallback

When no accessible, human-in-the-loop option exists for users affected by AI errors – forcing them into inaccessible loops of automation.

Failure to test for disability equity

When organizations treat accessibility as a UI issue only, and don't audit AI models for fair outcomes across disability groups.

These are not merely ethical lapses; they are functional barriers which are just as real as missing captions or unlabeled form controls.

The tester's pivotal role

Testers sit at the intersection of product reality and user experience, with a great vantage point to identify any exclusion. They can:

- **Expand test oracles:** Define expected outcomes for diverse disability contexts, not just UI correctness.
- **Introduce representative data:** Use curated datasets that reflect disability-related variation in speech, movement, timing, and AT interaction patterns.
- **Run fairness evaluations:** Compare performance across user segments, stratified by interaction mode (e.g., keyboard vs. mouse), speech profiles, or timing differences.
- **Stress-test explanations:** Validate that the reasons behind decisions are understandable, actionable, and accessible to the users who require any accommodations.
- **Advocate for process changes:** Embed accessibility checkpoints in model lifecycle reviews, not only in design QA.

Advanced testing methods for accessibility in AI

Ensuring that AI systems deliver accessible and equitable outcomes requires testing methods that go beyond conventional interface checks. Traditional accessibility evaluations, such as color contrast or screen reader compatibility, do not capture the deeper algorithmic biases that affect disabled users. To uncover these hidden barriers, some of the below techniques can be applied to evaluate fairness, consistency, and real-world resilience of AI models.

Metamorphic testing for accessibility

Metamorphic testing identifies bias or inconsistency by altering inputs in ways that should not change the output. In accessibility testing, this approach can be used to determine whether AI systems react fairly to disability-related variations. For example, a résumé-screening algorithm can be tested by introducing minor changes such as employment gaps due to medical leave or mention of assistive technology. If the model's score drops disproportionately, it indicates the presence of embedded bias. Similarly, adjusting voice or navigation inputs, like using screen readers or alternative interaction methods, should not trigger degraded responses or flag users as "suspicious."

Metamorphic testing thus validates whether AI outputs remain consistent and equitable when faced with legitimate differences in user interaction or background.

Adversarial testing for accessibility

Adversarial testing deliberately challenges AI models with complex, atypical, or edge-case inputs to expose discriminatory or inaccessible behavior. In accessibility contexts, these “edge cases” often mirror real experiences of disabled individuals. For instance, speech recognition systems can be tested using samples with stutters, dysarthria, or synthetic assistive voices to measure fairness in transcription accuracy. Likewise, facial or emotion recognition systems can be evaluated with images that include assistive devices or visible differences to reveal misclassification or exclusion.

By probing how AI reacts to disability-related diversity, adversarial testing highlights the fragility of models that were not trained on inclusive datasets, revealing where systems fail to accommodate human variation.

Error guessing for accessibility

Error guessing draws on the tester’s domain knowledge and intuition to anticipate where accessibility issues are most likely to occur. Testers familiar with disability use cases can identify high-risk scenarios that automated tools often miss. For example, they may predict that hiring systems penalize résumé gaps, that content moderation wrongly flags mobility aids as unsafe, or that timed authentication processes disadvantage users with slower motor responses.

This proactive approach directs testing efforts toward likely points of exclusion, ensuring that testing resources are focused where bias or inaccessibility is most harmful.

Explainability and transparency testing

This method tests whether AI explanations are understandable and accessible to users, including those using assistive technologies. Ensuring that automated decision explanations (like “why your application was declined”) are presented in plain language, readable formats, and compatible with screen readers.

Accessible transparency empowers users to question or appeal AI outcomes effectively.

Continuous monitoring and drift testing

This method tests whether AI explanations are understandable and accessible to users, including AI performance can degrade over time, especially for underrepresented user groups. Continuous monitoring detects such “bias drift” early. Tracking whether a speech recognition model’s accuracy for users with dysarthria declines after retraining or updates.

Regular audits maintain accessibility integrity throughout the AI lifecycle.

Key takeaways

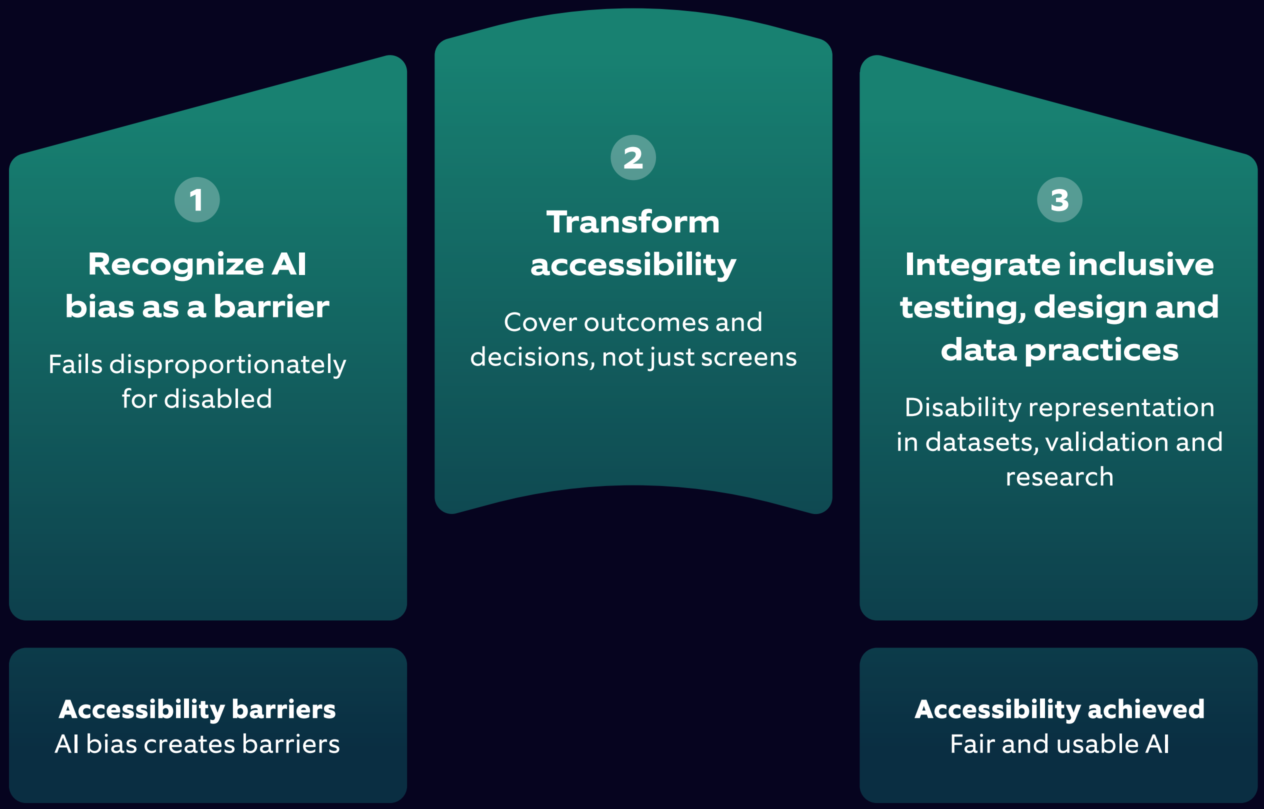

Figure 1: How AI bias creates accessibility barriers

- **Recognize how AI bias creates accessibility barriers:** If a decision system fails disproportionately for disabled users, that is an accessibility failure, regardless of interface polish.
- **Transform accessibility from design compliance to AI decision-making:** Accessibility must cover outcomes, not just screens. Evaluate fairness alongside usability.
- **Integrate inclusive testing, data, and design practices:** Bring disability representation into datasets, validation, and user research. Monitor, explain, and remediate issues across the entire lifecycle.

How to achieve these takeaways

1 Build representative datasets and tests

- Curate datasets that include disabled voices, interaction patterns, and assistive technology usage. Document coverage and known gaps.
- Establish segment-level metrics (error rates, false positives/negatives) across disability-relevant interaction modes.
- Use targeted synthetic data judiciously to augment coverage but validate with real users to prevent synthetic bias drift.

2 Adopt an accessibility-first model lifecycle

- **Problem framing:** Explicitly consider disability implications and potential proxies in the objective and constraints.
- **Data collection and labeling:** Include guidelines for annotators on disability contexts; audit label consistency and stereotypes.
- **Model development:** Optimize not just for global accuracy but for subgroup parity. Use techniques like reweighting, counterfactual evaluation, and constrained optimization where appropriate.
- **Pre-deployment evaluation:** Run structured fairness reviews that include disabled user scenarios, assistive tech interaction tests, and accessibility heuristics for explanations.
- **Deployment and monitoring:** Track performance by interaction mode and surface user-reported issues effectively. Establish red lines for rollback or human review.

3 Provide accessible alernatives and human-in-the-loop

- Offer multiple input/output modalities (voice, text, visual, haptic) with feature parity.
- Create clear, accessible appeal mechanisms with time-bound human review.
- Supply explanation summaries in plain language and compatible formats for assistive tech.

4 Institutionalize accountability

- **Define responsibility:** Assign owners for model fairness and accessibility at each stage.
- **Documentation:** Maintain model cards, data statements, and accessibility impact assessments that explicitly address disability.
- **Training and culture:** Equip teams with training on disability, inclusive research, and AI fairness. Include disabled experts and users in governance.

5 Test end-to-end journey

- Offer multiple input/output modalities (voice, text, visual, haptic) with feature parity.
- Create clear, accessible appeal mechanisms with time-bound human review.
- Supply explanation summaries in plain language and compatible formats for assistive tech.

Figure 2: Steps to achieve the takeaway goals

Other possible initiatives

- **Procure responsibly:** Require vendors to show bias and accessibility testing results.
- **Engage communities:** Partner with disability organizations for dataset annotation and review.
- **Advance standards:** Extend WCAG-style accessibility thinking to algorithmic decisions.
- **Invest in research:** Support studies on atypical speech recognition and inclusive AI.
- **Embrace personalization with safeguards:** Allow adaptive systems with privacy and opt-out options.

Conclusion

AI promised to democratize intelligence and expand opportunities at unprecedented scale. But that promise rings hollow when the systems designed to empower us systematically exclude disabled people. This isn‘t innovation stumbling toward progress; this is exclusion by algorithm.

We already recognize inaccessible websites as failures - legally, morally, and technically. Biased AI that blocks disabled people from jobs, services, and participation deserves the same judgment. The distinction we draw between physical barriers and algorithmic ones is arbitrary and untenable. Both are accessibility failures. Both demand the same urgency.

The reality is clear: accessibility cannot remain confined to interface design, to screen readers, and color contrast ratios only. It should also extend into the foundations of AI itself: the datasets that train models, the logic that drives decisions, and the outcomes that shape lives. Surface-level compliance is no longer sufficient when the machinery of intelligence operates beneath that surface.

This solution is not only aspirational but also mandatory. It should involve representative datasets that capture human diversity and inclusive testing that comprises disabled users from the very beginning. It would also have transparent reasoning that can be examined and challenged, with accessible alternatives when systems fail, and continuous human oversight that never disappears. All these aren‘t just nice-to-haves. They‘re prerequisites for responsible AI.

We stand at a crossroads. One path leads to intelligence at scale but without inclusion at scale. It has technology that works brilliantly for most people while leaving millions behind. The other path requires us to rebuild AI‘s architecture around accessibility from the ground up, ensuring that expanded capability means expanded access for everyone.

The choice will define whether AI fulfills its promise or betrays it. The technology is powerful enough to reshape society. The question is whether we‘re wise enough to ensure it reshapes society justly – not just for most of us, but for **all of us.**

About the author

Anamika Mukhopadhyay

Global Practice Lead – Special & Emerging Tech Testing, Nagarro

Anamika Mukhopadhyay is a seasoned QA Consultant with a decade-long experience and a keen understanding of how functionality, performance, user experience, and accessibility intricately intertwine. She guides enterprises in setting up top-notch test automation capabilities. Her dynamic perspective fosters innovative testing strategies and her passion lies in creating inclusive software that caters to diverse user needs. When not at work, she enjoys exploring new cities and savoring the local cuisine.

anamika.mukhopadhyay@nagarro.com

About Nagarro

Nagarro helps future-proof your business through a forward-thinking, fluidic, and CARING mindset. We excel at digital engineering and help our clients become human-centric, digital-first organizations, augmenting their ability to be responsive, efficient, intimate, creative, and sustainable. Today, we are around 17,500 experts across 39 countries, forming a Nation of Nagarrians, ready to help our customers succeed.

For more information, visit **www.nagarro.com**.