IIAR JOURNALS

Innovating Knowledge. Advancing Research.

IIAR Journal of Disability, Inclusion, and Assistive Technologies

Vol 1, No 1, 2026

Algorithmic Exclusion: A Conceptual Model of How Artificial Intelligence Systems Disable, from Training Data to Deployment

IGWE Emmanuel Sunday[1], NWOGBOJI Emmanuel Usulor[2], OGAH Sunday Ogbo[3], BAMIGBOLA Paul Idowu[4], ATOLAGBE Umar Kawu[5]

[1,3,4,5] Veristzon Global Research Lab; [2] International Institute of AI and Research

ABSTRACT

The literature on fairness in artificial intelligence has organised itself around race and gender. Disability, by contrast, is rarely treated as a fairness category at all, even though disabled people are among the groups most exposed to algorithmic harm: speech recognisers that cannot transcribe dysarthric speech, emotion classifiers built on a norm of expressiveness that autistic people do not share, hiring filters that penalise employment gaps, and proctoring software built on assumptions about how a test-taker's body and circumstances should look. This conceptual paper asks how AI systems disable, and why the standard fairness toolkit answers so poorly when they do. Drawing on the social model of disability and critical disability studies, on the algorithmic-fairness scholarship and its legal and feminist critiques, and on peer-reviewed research published between 2023 and 2026 on the performance of speech, facial, visual, language, hiring and screening systems for disabled users, the paper develops a four-mechanism model of algorithmic exclusion: under-representation in training data, normative modelling that treats atypical bodies and behaviours as noise or risk, inaccessible interfaces to the system itself, and deployment contexts that offer no human alternative. It then argues that disability defeats group-based fairness remedies for four structural reasons, namely heterogeneity, invisibility, legally protected non-disclosure and poor fit with the categorical variables on which parity metrics depend. Eight propositions locate where exclusion is most probable and most damaging, with particular attention to African settings where disability data are scarcest and where regulatory capacity to demand accessible AI is thinnest. The paper closes with a design-and-procurement framework that treats disability as a first-order constraint rather than an edge case, and with recommendations for developers, public purchasers, data-protection authorities, organisations of persons with disabilities and the speech-technology and fairness research communities.

Keywords: algorithmic bias; disability; accessibility; fairness; speech recognition

Algorithmic Exclusion: A Conceptual Model of How Artificial Intelligence Systems Disable, from Training Data to Deployment

1. INTRODUCTION

When the fairness of an artificial intelligence system is audited, the auditor almost always reaches for the same pair of variables. Does the system perform equally well for men and women? Does it treat racial groups alike? These are the right questions, but they are not the only ones, and the habit of asking only these has produced a literature that is far more confident about some forms of harm than about others. Disability is the clearest casualty of that habit. The most comprehensive survey of bias in large language models catalogues evaluation datasets by the social group they target, and disability appears there as one column among many rather than as an organising concern (Gallegos et al., 2024). A scoping review of a decade of health-related AI research for disabled people found that fewer than one study in fifteen involved disabled people through co-design or user validation, and only one in eight implemented any concrete equity measure (Kaboré et al., 2026). The field has been building systems for people it has rarely consulted and auditing them for harms it has rarely measured.

The practical consequences are now well documented, even if they remain scattered across disciplines that do not read one another. A baseline speech recogniser that transcribes typical speech with a 3.4 per cent word error rate produces a 36.3 per cent error rate for speakers with dysarthria (Hasegawa-Johnson et al., 2024). Commercial recognisers achieve at best two-thirds word accuracy for young adults with Down syndrome (Cibrian et al., 2025). In the visual datasets used to train perception systems for autonomous vehicles, wheelchair users make up less than 1 per cent of the people annotated (Fernández Llorca et al., 2024). Large language models, asked to write stories about people with intellectual disabilities, reliably make them younger, more dependent and more in need of rescue than they make everyone else (Coffey et al., 2026). None of these findings is surprising to disabled people, who have been reporting them for years. What is new is that they can now be assembled into an argument.

That argument is the purpose of this paper. I want to move the discussion from a list of failures to a model of how the failures are produced, because only a model allows one to predict where the next failure will appear and to design against it. The paper makes three claims. The first is that AI systems disable through four distinct mechanisms, operating at different points between data collection and deployment and calling for different remedies: under-representation in training data, normative modelling that treats atypical bodies and behaviours as noise or risk, inaccessible interfaces to the system itself, and deployment contexts that offer no human alternative. The second is that disability is not one more protected attribute to be added to the fairness checklist; it poses a structural problem for the group-parity paradigm because disability is heterogeneous, frequently invisible, legally shielded from disclosure and poorly captured by the categorical variables on which fairness metrics depend. The third is that the remedy therefore lies less in better metrics than in treating disability as a first-order design and procurement constraint, enforced where organisations decide what to buy.

The paper is written with African conditions in view. Disability statistics on the continent are patchy, the Washington Group questions that underpin most of them yield different estimates depending on the cut-off chosen (Hanass-Hancock et al., 2023), and the data protection regimes of even the better-resourced jurisdictions, Kenya and South Africa, do not yet adequately protect disabled people from algorithmic discrimination, neither treating disability data as a special category (Genga, 2025). If algorithmic exclusion is a function of data scarcity and weak regulatory reach, Africa is where the model should predict it will bite hardest, and the paper states that prediction explicitly, as a claim that empirical work could refute, rather than treating the continent as an afterthought. Section 2 reviews the concepts, theories and evidence; Section 3 describes the method of synthesis; Section 4 presents and discusses the model, the propositions and a design-and-procurement framework; Sections 5 and 6 conclude and state the contribution.

2. LITERATURE REVIEW

2.1 CONCEPTUAL REVIEW

Three terms carry most of the weight in what follows and each needs to be fixed before use. *Disability* is understood here in the sense established by the social model and carried into the Convention on the Rights of Persons with Disabilities: the restriction of participation that results when a person with an impairment meets an environment, including a

technical environment, built without that impairment in mind. Adam and Koutsoklenis (2023) have recently defended the social model against three lines of criticism, arguing that it remains indispensable for disabled people who have been denied their dignity in material and cultural terms; what makes the model useful for thinking about software is that same refusal to let the environment off the hook. A speech recogniser that cannot understand dysarthric speech does not reveal a deficit in the speaker. It reveals a decision, usually unexamined, about whose speech counted when the system was built.

Algorithmic exclusion is used in preference to the more familiar *algorithmic bias* for a deliberate reason. Bias suggests a measurable deviation from a correct value, something that can be reduced by better estimation. Exclusion names an outcome: a person who cannot use, cannot be seen by, or cannot contest a system. The distinction matters because some of the most consequential harms to disabled people are not biases in any statistical sense. An interface that a screen-reader user cannot operate is not biased; it is closed. A deployment that removes the human clerk is not biased; it removes the only route by which an atypical case could ever have been handled. Foley and Melese (2026) make a related move when they write of *disabling* AI, showing through case analyses of chatbots, learning analytics and tutoring systems how educational AI embeds ableist assumptions and sidelines non-normative ways of learning and knowing; the model developed here takes that verb seriously, as an action performed on people rather than a property of them.

Accessibility and *fairness* are often run together, but they are different goods and are pursued by different communities with different tools. Accessibility asks whether a person can operate a system; fairness asks whether the system's outputs are distributed justly. A hiring algorithm can be perfectly accessible to a blind applicant and still rank her unfairly; a proctoring system can be fair in its false-positive rate across groups and still be unusable by a student who cannot hold her gaze on a screen. Much of the confusion in the literature comes from satisfying one and claiming the other. Barton et al. (2025) note that accessibility standards such as the Web Content Accessibility Guidelines have had suboptimal impact because of limited comprehensiveness and weak enforcement mechanisms, and that they rarely account for variation within the disability community; the gap between the two vocabularies is one of the places algorithmic exclusion lives.

2.2 THEORETICAL REVIEW

The theoretical spine of the paper is the claim, developed most carefully by Newman-Griffis et al. (2023), that definition drives design. The model of disability a development team holds, usually without articulating it, shapes how the problem is framed, which data are treated as relevant, what the system is used for and how it is operated, and each of these choices can look neutral on its own. A team working from a medical model builds systems that detect and classify impairment; a team working from a social model builds systems that remove barriers; a team working from no model at all builds systems that treat the statistically dominant body as the body. The last is by far the most common, and it is the source of what İnce (2026) calls algorithmic ableism: the reproduction, in data, models and everyday design discourse alike, of assumptions about what counts as a normal body, sense or form of communication, such that the atypical body registers as something to be converted into a format the system can read. Moura (2023) traces the same logic through two cases, the allocation of lifesaving medical interventions during the COVID-19 pandemic and approaches to autism diagnosis and intervention, and shows that particular ethical norms can become part of technological systems, producing a mismatch between disabled people's lived experience and the goals of the non-disabled people who build them.

A second theoretical resource is the critical literature on fairness itself. The technical paradigm defines fairness as parity of some statistic, such as error rate or selection rate, across groups defined by a protected attribute (Gallegos et al., 2024), and the policy advice that has grown up around it, including the best practices that Alvarez et al. (2024) set out within a legal layer focused on the European Union and a bias-management layer, is built on that framing. Legal scholars have argued that technical fairness strategies are misaligned with anti-discrimination and data-protection law, because technically "fair" models neglect individualised harm and proportionality and common mitigation methods can erase the evidence of structural discrimination that a legal claim would need (Liu et al., 2026), and feminist scholars have argued that it cannot be squared with intersectional reality, because the groups it compares are narrow and the structures it ignores are wide (Mirsch et al., 2026). I draw on both critiques, but I add a point that neither makes: the group-parity paradigm presupposes a group, and disability frequently does not supply one. The theoretical consequence is that the standard remedies, which are remedies for group disparity, have nothing to grip.

The third resource is the sociotechnical analysis of AI and independent living that Mladenov (2026) has developed from within disability studies. Mladenov rejects both techno-hype and techno-anxiety, treating AI as something that can be cure or poison for independent living depending on how it is governed and whether disabled people shape it, and he identifies the overvaluation of self-sufficiency and what he calls techno-fetishism, the opacity that hides human agency behind the system, as specific dangers. This framing explains why the fourth mechanism in my model, the removal of the human alternative, is so damaging: it converts a system that could have been a tool into an obligatory passage point. Umucu (2025) supplies a complementary psychological account, a framework for disability-inclusive AI design grounded in self-determination theory and self-efficacy theory, which likewise treats the person's own agency as the thing the technology must serve rather than replace.

2.3 EMPIRICAL REVIEW

The empirical record is uneven, and I report it domain by domain because the mechanisms differ.

Speech is the best-studied domain and the clearest case of training-data exclusion. The Speech Accessibility Project, funded by a coalition of five technology companies and run by a university, was established because recognition of dysarthric speech had improved far more slowly than recognition of typical speech, for want of large, shareable corpora; its first data release, 211 speakers with dysarthria associated with Parkinson's disease, documented a baseline word error rate of 36.3 per cent for dysarthric speakers against 3.4 per cent for typical speakers, falling to 23.7 per cent after fine-tuning (Hasegawa-Johnson et al., 2024). That improvement is real but partial. Speaker-independent models fine-tuned on the TORGO and UA-Speech corpora improve markedly on those corpora yet, tested across datasets, tend to collapse continuous severe dysarthric speech into isolated words (Zhong et al., 2025). Six commercial recognisers tested on 15 young adults with Down syndrome achieved at best 67 per cent word accuracy (Cibrian et al., 2025). Sohrabi (2026) adds an observation that matters for Africa: the failure patterns of recognisers on impaired speech, including hallucinated content and language confusion, are structurally identical to their failure patterns on Persian, a low-resource language, which Sohrabi attributes to a shared mechanism of distributional exclusion rather than to anything specific to impairment. Co-design with people with dysarthria confirms that the fix is not better recognition alone but a rethinking of the interaction (Pucci et al., 2024).

Facial and affective computing illustrates normative modelling rather than data scarcity. Mattioli and Cabitza (2024) review the evidence against the universality hypothesis on which automatic facial emotion recognition rests and find systems built on limited datasets that ignore context. McInerney and Keyes (2025) go further, arguing from the industry's own self-presentation that emotion recognition technology encodes a norm of legible expressiveness against which the autistic person is positioned as deficient and in need of correction. The point is not that these systems lack autistic training data, though they may; it is that the construct they measure is defined so that autistic expression cannot score well.

Computer vision for the physical world shows the first mechanism in its purest form. Fernández Llorca et al. (2024) annotated more than 90,000 people across six visual datasets used to train perception systems for autonomous driving and found wheelchair and personal mobility device users to be extremely under-represented, at less than 1 per cent of samples, with diversity very low for most attributes; the evaluation of model performance for these groups is left to future work, and the data make it plain why it will be hard.

Language models present a representational harm. Coffey et al. (2026) generated 25,000 stories across five models and found that an intellectual-disability descriptor shifted the narrative towards youth, paternalism, dependence and rescue. Mutia et al. (2026) found ableist classifications in a substantial share of responses from three commercial chatbots in English and Bahasa Indonesia, concentrated where disability intersected with lower socioeconomic status. At the same time, disabled students report real benefit from these models for reading, writing and self-organisation (Pierrès et al., 2025); exclusion and assistance coexist in the same artefact.

Hiring is where exclusion meets consequence. Fabris et al. (2025) survey the multidisciplinary evidence on algorithmic hiring and show how much of it concerns gender and race, with discrimination against disabled applicants less studied even though the signals these systems rely on, including gaps in employment history, speech and eye contact in asynchronous video interviews and gameplay in psychometric assessment, can correlate with impairment. Rigotti and Fosch-Villaronga (2024) examine the same systems from a legal direction and find that existing legal frameworks apply to algorithmic recruitment

but address its fairness only piecemeal. The empirical thinness here should be stated plainly: there is little published audit evidence on how specific commercial hiring tools treat specific disabilities, partly because the tools are proprietary and partly because applicants do not know they were screened.

Proctoring and welfare screening illustrate the deployment mechanism. Selwyn et al. (2023) document the rapid adoption of online exam proctoring in Australian universities during the first pandemic year and the concerns it raised about the surrender of control to commercial providers, the hidden labour behind supposedly automated systems and the increased vulnerabilities of remote study; what such systems assume about a test-taker's body and circumstances is my inference from how they work, and the peer-reviewed literature on proctoring and disability specifically remains thin. Moura et al. (2026) evaluate disability bias in the Allegheny Family Screening Tool and show that, although the tool contains no explicit disability measure, it captures disability through proxies such as healthcare use, public benefits and contact with the criminal legal system, so that disabled parents and children would receive worse risk scores than their non-disabled peers. The screening case therefore shows two mechanisms at once: the risk model treats disability-related circumstance as risk, which is normative modelling, and the score is acted on at scale, which is deployment. In both the proctoring and the screening case the harm is amplified by the absence of a human route around the system. Interfaces to AI systems are a further barrier. Esposito et al. (2026) compared a conversational-AI approach to web browsing with screen readers among 30 blind and low-vision users, and the fact that such a comparison was needed makes the point: the accessibility of a conversational interface is an empirical question to be settled with these users, not a property of the modality.

Finally, the African evidence. Dalvit (2025) examines language use, algorithmic bias and AI-generated content in South African online representations of disability and suffering and argues that they reproduce existing tropes, the helpless victim, the resentful villain and the inspiring hero. Genga (2025) concludes that, despite progress, neither Kenya's nor South Africa's data protection framework yet adequately protects disabled people from algorithmic discrimination. Plantinga et al. (2024) find, across twelve African countries, that AI strategies and ethics frameworks are often drafted without assessing their fit with local context or their implementability, while existing governance capabilities in banking model management, air-quality assessment and electronic-transactions law that could inform responsible AI go largely unexamined. On the data side, the Disability Statistics Estimates Database now provides comparable indicators for forty countries (Carpenter et al., 2024), but the underlying Washington Group measures yield prevalence estimates that depend on which cut-off on the degree of functional difficulty is applied, a choice still under discussion (Hanass-Hancock et al., 2023), and even in East Africa, where survey coverage is comparatively good, reported disability prevalence ranges from about 10 per cent in the Democratic Republic of Congo to about 20 per cent in Kenya and is higher in the poorest households (Luoga et al., 2026). The data on which any disability-aware AI would have to be built are contested at the point of definition.

3. MATERIALS AND METHODS

This is a conceptual paper, and its method is theoretical synthesis rather than empirical measurement. The design has three steps: a structured review of recent peer-reviewed evidence, the abductive derivation of mechanisms that account for the patterns in that evidence, and the development of propositions that could, in principle, be falsified by future empirical work.

The literature was retrieved between September and October 2026. Candidate articles were identified through PubMed and the OpenAlex scholarly index, supplemented by general web searching for articles in specific journals. Searches combined disability terms (disability, disabled, dysarthria, Down syndrome, autism, intellectual disability, wheelchair, blind, deaf) with system terms (artificial intelligence, machine learning, algorithm, speech recognition, facial recognition, emotion recognition, large language model, hiring, proctoring, screening) and outcome terms (bias, fairness, ableism, exclusion, accessibility). Inclusion required that an item be a peer-reviewed journal article with a registered DOI, published or issued in 2023 to 2026, and that it bear on at least one of the four mechanisms or on the African context. Preprints, conference proceedings, book chapters, reports and grey literature were excluded even where influential, so several well-known studies presented at computing conferences are not cited; this is a limitation of the evidence base rather than a judgement on those studies. Approximately 180 records were screened at title and

abstract level, 52 were examined in full, and 39 are cited.

Every cited reference was verified against its Crossref publisher record, or where Crossref was temporarily unavailable against the publisher's own landing page, for author names, title, journal, year, volume, issue and pagination, and each record was checked for retraction, correction or expression-of-concern relations. None was found. Bibliographic details in the reference list are taken from those records rather than from secondary citations.

Mechanisms were derived abductively: for each documented failure, I asked at what point in the pipeline it was introduced and what would have had to be different to prevent it, grouped failures that shared an answer, and tested the groups against new cases until none required a fifth group. Propositions were then written as conditional statements linking features of the data, the model, the interface or the deployment to the probability and severity of exclusion. The basis of judgement throughout is coherence with the evidence and with the social model of disability; where the evidence was thin, as it is for proctoring and for Africa, I say so rather than extrapolate.

No human participants were involved and no ethical approval was required. There is no dataset beyond the reference list, which is openly available. An AI writing assistant was used, under the author's direction, to support literature retrieval, bibliographic verification, drafting and formatting; the author is responsible for every claim, interpretation and citation in the paper.

4. RESULTS AND DISCUSSION

4.1 RESULTS

The synthesis yields three results: a four-mechanism model, an account of why disability resists the standard fairness remedies, and propositions with an accompanying design-and-procurement framework.

The four mechanisms are set out in Table 1, distinguished by where in the pipeline they arise, by the form the exclusion takes and by the remedy each demands. Under-representation is the mechanism the fairness literature already understands, and it explains the speech and vision findings: systems trained on bodies and voices from which disabled people were absent cannot recognise them. Normative modelling is subtler and explains the affective-computing, language-model and screening findings: the construct the system measures, or the narrative it has learned to produce, is defined so that atypical expression or circumstance scores as deficit, dependence or risk, and adding data does not help because the data would be labelled by the same norm. The inaccessible interface is the oldest problem in accessibility wearing new clothes: the system may be fair in its outputs and still closed to the person who cannot operate its input. Deployment without alternative is a property of the organisation rather than the software, and it converts each of the other three from an inconvenience into a wall.

Table. 1. Four mechanisms of algorithmic exclusion

Mechanism	Pipeline stage	What is excluded	Characteristic failure	Illustrative evidence	Primary remedy
Under-representation	Data collection and curation	Atypical bodies, voices and movements absent from training data	System cannot recognise or transcribe the person	Hasegawa-Johnson et al. (2024); Cibrian et al. (2025); Fernández Llorca et al. (2024)	Purposive data collection led by disabled communities; reporting of performance by impairment
Normative modelling	Problem definition, labelling and feature selection	Atypical expression, behaviour or circumstance treated as noise, deficit or risk	System recognises the person but misclassifies or penalises them	McInerney and Keyes (2025); Moura (2023); Moura et al. (2026); Fabris et al. (2025); Coffey et al. (2026); Mutia et al. (2026)	Re-examination of the construct and its labels; removal of impairment-correlated proxies
Inaccessible interface	Interaction design	Users who cannot operate the input or perceive the output	Person cannot use, contest or even know of the system	Esposito et al. (2026); Barton et al. (2025)	Accessibility conformance as a release gate; multimodal interaction
Deployment without alternative	Organisational adoption and procurement	The human route around the system	Any of the above becomes decisive rather than	Selwyn et al. (2023); Mladenov (2026); Moura et al.	Guaranteed human alternative; procurement

			inconvenient	(2026)	conditions; appeal rights

The second result concerns why disability is a poor fit for group-parity remedies. Four features of disability combine to defeat them. Disability is *heterogeneous*: a wheelchair user, a person with dysarthria and an autistic person share no feature that a model could learn, and a parity metric computed over "disabled" as a single group would average their very different experiences into a figure that describes none of them. Disability is frequently *invisible*: a sizeable proportion of the working population has a disability that is not visible, a share that is growing with the prevalence of mental ill health, and many choose not to disclose it (Hassard et al., 2024), so such impairments leave no trace in the data a system sees and the group cannot be identified even in principle. Disability is *protected from disclosure*: law in many jurisdictions prohibits employers and service providers from asking, and the wording of even voluntary disclosure forms changes who answers (Santuzzi et al., 2025), so the ground-truth labels that fairness auditing requires are unobtainable without violating the very protections the audit is meant to serve. And disability is *poorly captured by categorical variables*: fairness metrics presuppose a small number of discrete groups, whereas the Washington Group questions grade functional difficulty by degree across several domains, so that who counts as disabled depends on the cut-off applied (Hanass-Hancock et al., 2023). The fairness literature in medicine has not resolved this: a recent scoping review of fairness metrics for clinical prediction models reports conceptual challenges in defining and quantifying fairness and a gap on intersectionality (Matos et al., 2026), and the assessments stratified across subpopulations on which Chen et al. (2023) rely presuppose that the subpopulations can be defined. A fairness audit can therefore report parity across every group it can see and still miss every disabled person the system has excluded.

The propositions in Table 2 follow from the model and from these four features. They are written as conditional claims about where exclusion is most probable and most harmful.

Table. 2. Propositions on the location and severity of algorithmic exclusion

Proposition	Statement	Mechanism	Supporting literature
P1	The probability of exclusion by under-representation rises with the rarity and the within-group variability of the impairment; it is highest where impairment is both uncommon and heterogeneous in presentation.	Under-representation	Zhong et al. (2025); Cibrian et al. (2025); Fernández Llorca et al. (2024)
P2	Exclusion by under-representation is compounded by low-resource language or dialect, so that disabled speakers of African languages face a multiplicative rather than additive penalty.	Under-representation	Sohrabi (2026); Hasegawa-Johnson et al. (2024)
P3	Where the construct a system measures is defined by reference to normative expressiveness, behaviour or life course, adding disabled training data will not remove exclusion, because the labels encode the norm.	Normative modelling	McInerney and Keyes (2025); Mattioli and Cabitza (2024); Newman-Griffis et al. (2023)
P4	Systems that use proxies correlated with impairment (employment continuity, gaze stability, response latency, service-use history) will penalise disabled people even when disability is never recorded, and group-parity audits will not detect it.	Normative modelling	Fabris et al. (2025); Moura et al. (2026); Hassard et al. (2024)
P5	Conversational and voice interfaces marketed as	Inaccessible interface	Esposito et al. (2026); Pucci et al. (2024); İnce (2026)

	accessible alternatives reproduce exclusion when their interaction patterns assume typical speech, sight or cognition; their accessibility must be demonstrated with disabled users rather than inferred from the modality.		
P6	The severity of exclusion is determined less by the model's error rate than by whether a human alternative exists; the same error is an inconvenience where a clerk remains and a denial where the clerk has been removed.	Deployment without alternative	Mladenov (2026); Selwyn et al. (2023); Moura et al. (2026)
P7	Representational harms in generative systems are concentrated at the intersection of disability with poverty and with intellectual impairment, where the dominant narrative of dependence and rescue is reproduced most reliably.	Normative modelling	Coffey et al. (2026); Mutia et al. (2026)
P8	In jurisdictions where disability data are sparse and data-protection law is silent on algorithmic discrimination, all four mechanisms operate with less friction, so exclusion is more probable, less visible and less remediable.	All four	Genga (2025); Plantinga et al. (2024); Carpenter et al. (2024); Kaboré et al. (2026)

The third result is the framework in Table 3. Its organising idea is that if group-parity auditing cannot protect disabled people, protection must move upstream to the decisions that determine what is built and bought. The framework assigns responsibilities by actor and by mechanism, and it treats procurement as the central instrument because public purchasers in most African countries have more influence over vendors than regulators do.

Table. 3. A design-and-procurement framework for disability-inclusive AI

Mechanism addressed	**Design requirement (developer)**	**Procurement condition (purchaser)**	**Verification**	**Role of organisations of persons with disabilities**
Under-representation	Performance reported by impairment type and severity; disabled people participate in data collection	Vendor discloses disaggregated performance and provenance of disability data; absence of data is disclosed, not assumed benign	Independent testing with disabled users before award	Co-lead data collection; set the categories to be reported
Normative modelling	Construct validity review: whose expression defines the norm? Removal of impairment-correlated proxies	Documentation of features and justification of any proxy that could correlate with impairment	Feature audit by a disability-competent reviewer	Identify proxies from lived experience that auditors miss
Inaccessible interface	Accessibility conformance as a release gate, extended to conversational and voice modalities	Conformance is a contract term with remedies	Testing with assistive technology users	Conduct usability testing; certify or refuse
Deployment without alternative	Design for contestability: decisions can be explained, appealed and overridden	A human alternative of equal timeliness is a condition of deployment	Reporting on appeals and overrides, where lawfully obtainable	Monitor the human route; raise systemic failures

4.2 DISCUSSION

The model's first claim is that the four mechanisms are distinct, and the evidence supports this in the sense that remedies appropriate to one mechanism demonstrably fail against another. The Speech Accessibility Project is a remedy for under-representation, and it works as such: fine-tuning on the new corpus cut the word error rate by a third (Hasegawa-Johnson et al., 2024). But the cross-dataset collapse reported by Zhong et al. (2025) shows that more data of the same kind does not generalise to speakers unlike those collected, which is P1 in operation, and nothing in the project touches the fact that a transcription service deployed as the only channel to a government office still excludes the speaker it mis-transcribes, which is P6. The affective-computing case shows the converse: data are not the problem at all. One could collect a million autistic faces and the emotion classifier would still score them as deficient, because the construct of a legible emotion was defined by a norm of expressiveness that autistic people do not share (McInerney & Keyes, 2025). Treating this as a data problem would waste effort and, worse, legitimate the construct.

The second claim, that disability resists group-parity remedies, is where the paper departs most sharply from the fairness mainstream, so the strongest objection should be met. Other protected attributes are also heterogeneous, sometimes invisible and sometimes protected from collection, and the fairness community has developed proxy-inference and fairness-without-demographics techniques for exactly these situations. The reply is that disability combines all four features at once and to a degree that other attributes do not. Race and gender are heterogeneous but share enough structure for a model to learn; disability does not. Race and gender are sometimes invisible in data but usually inferable from it; a psychiatric or cognitive impairment often is not. And the legal protection of disability non-disclosure is not a nuisance to be worked around by inference but a right, so that a fairness technique which infers disability in order to audit for it has already committed the harm it was meant to detect. Mirsch et al. (2026) argue that substantive fairness may sometimes require non-deployment; the argument here is narrower, that it requires moving the point of intervention from the audit to the specification.

Some propositions are better supported than others. P1, P3, P4 and P7 rest on direct evidence from several studies. P2, the multiplicative penalty for disabled speakers of low-resource languages, rests on Sohrabi's (2026) convergence finding and on inference; no study I could find has tested dysarthric speech in an African language against a commercial recogniser, and this is the single most important empirical gap the model identifies. P5 rests on two studies and on İnce's (2026) account of algorithmic ableism, which supplies the reason a modality cannot be assumed accessible: the normative body is encoded in the interaction pattern as much as in the model. P6 has the strongest theoretical grounding and the weakest direct measurement, because the counterfactual, what would have happened had a human been available, is rarely recorded. P8 is supported by legal and policy analysis rather than outcome data, for the reason the proposition itself gives: where data are sparse, exclusion is invisible.

The third claim, that the remedy lies in specification and procurement rather than in better metrics, is tested most severely by P8, and it would be easy to read P8 as a counsel of despair. It is not. The finding of Plantinga et al. (2024) that African AI governance leaves existing sectoral mechanisms largely unexamined is also a finding that such mechanisms exist, in the model management procedures of banking and in electronic-transactions law, to which I would add public procurement law and the disability legislation most states have enacted since ratifying the Convention. The framework in Table 3 is designed to run through those mechanisms rather than to require new ones; a procurement condition requiring a human alternative needs no AI-specific statute, only a procurement officer who has been told to write it. This is also why the framework gives organisations of persons with disabilities operational rather than consultative roles. Kaboré et al. (2026) found that only six of ninety-three AI studies involved disabled people in design; the remedy for that figure is not a seat at an advisory table but authority over the categories by which performance is reported and the tests by which conformance is judged.

There are limits to what a conceptual paper can establish. The mechanisms are derived from a published record that over-represents English, high-income settings and the domains that attract computing researchers; a fifth mechanism might emerge from systematic study of, say, agricultural or financial AI in rural Africa. The exclusion of conference proceedings removed several influential studies and may have tilted the review towards domains with established journals. And the paper has treated disability largely in isolation from gender,

age and poverty, whereas Mutia et al. (2026) show that ableist output concentrates where disability intersects with lower socioeconomic class, and Dalvit (2025) places South African online representations of disability against the country's profound social inequalities. Each of these is a direction for empirical work, and the propositions have been written so that such work could refute them.

5. CONCLUSION AND RECOMMENDATION

This paper has argued that artificial intelligence systems disable through four mechanisms that arise at different points between the collection of training data and the organisational decision to deploy, that disability poses a structural rather than an incidental problem for the group-parity paradigm of algorithmic fairness, and that protection must therefore move upstream from the audit to the specification and the purchase order. The four-mechanism model explains why some remedies work and others do not, the four-feature account explains why fairness audits can be clean while disabled people are excluded, and the eight propositions say where the next exclusions should be expected. The African framing is not decorative: it is the setting in which the model predicts the most exclusion and the least evidence of it, and also the setting in which existing procurement and disability law offer the most practical purchase.

The recommendations follow from Table 3 and are addressed to named actors. Developers of speech, vision and language systems should report performance disaggregated by impairment type and severity as a matter of routine, should subject every construct their systems claim to measure to a review of whose expression defines its norm, and should treat accessibility conformance, extended to conversational and voice modalities, as a release gate rather than a roadmap item. Public purchasers, including ministries of education, labour and social development and the agencies that run national identity and welfare systems, should write four conditions into every contract for an AI system that touches eligibility, assessment or access: disclosure of disaggregated performance, justification of any feature that could correlate with impairment, accessibility conformance as a contract term carrying remedies, and a guaranteed human alternative of equal timeliness. Data-protection authorities in Kenya, South Africa and their peers across the continent should issue guidance naming algorithmic discrimination against disabled people as unlawful processing and clarifying that inferring disability for the purpose of auditing is itself processing of special-category data requiring a lawful basis. Organisations of persons with disabilities should seek, and funders should resource, the technical capacity to set reporting categories, run conformance tests and monitor the human route, because consultation without authority changes little. Speech-technology researchers working in Africa should treat the testing of commercial recognisers on dysarthric speech in African languages as an urgent and tractable study. And the fairness research community should stop treating disability as an attribute to be appended to existing benchmarks and ask instead, for each benchmark, whether the harms it is built to catch are the harms disabled people experience.

6. CONTRIBUTION TO EXISTING KNOWLEDGE

The paper's first contribution is the four-mechanism model itself. Earlier work on AI and disability has either catalogued failures domain by domain or argued at a high level that AI reproduces ableism; the model sits between these, offering a small number of mechanisms distinguished by pipeline stage, form of exclusion and remedy, and shown to explain findings from speech, vision, affective computing, language models, hiring, proctoring and welfare screening. Because the mechanisms are tied to remedies, the model is usable: a practitioner who can locate a failure in Table 1 knows what to try and what not to bother with.

The second contribution is the argument that disability resists group-parity fairness remedies for four structural reasons. The legal and feminist critiques of the fairness paradigm have established that parity metrics are insufficient in general; this paper shows that for disability they can be inapplicable in principle, because the group the metric presupposes may not be identifiable without violating the rights the metric is meant to protect. That relocates the problem from measurement to specification and procurement, which is the third contribution: a framework assigning design requirements, procurement conditions, verification steps and operational roles for organisations of disabled people to each mechanism, built to run through institutions African states already possess. Finally, the paper contributes testable propositions and an explicit account of where the evidence is thin; the identification of dysarthric speech in African languages as an untested case, and of the human-

alternative counterfactual as an unmeasured one, gives empirical researchers a concrete agenda.

7. REFERENCES

Adam, S., & Koutsoklenis, A. (2023). Who needs the social model of disability? *Frontiers in Sociology, 8*, Article 1305301. https://doi.org/10.3389/fsoc.2023.1305301

Alvarez, J. M., Bringas Colmenarejo, A., Elobaid, A., Fabbrizzi, S., Fahimi, M., Ferrara, A., Ghodsi, S., Mougan, C., Papageorgiou, I., Reyero, P., Russo, M., Scott, K. M., State, L., Zhao, X., & Ruggieri, S. (2024). Policy advice and best practices on bias and fairness in AI. *Ethics and Information Technology, 26*(2), Article 31. https://doi.org/10.1007/s10676-024-09746-w

Barton, H. J., Valdez, R. S., Shew, A., Swenor, B. K., Jolliff, A., Claypool, H., Czaja, S. J., & Werner, N. E. (2025). A call for integrated approaches in digital technology design for aging and disability. *The Gerontologist, 65*(6), Article gnaf113. https://doi.org/10.1093/geront/gnaf113

Carpenter, B., Kamalakannan, S., Patchaiappan, K., Theiss, K., Yap, J., Hanass-Hancock, J., Murthy, G. V. S., Pinilla-Roncancio, M., Rivas Velarde, M., & Mitra, S. (2024). Data resource profile: The Disability Statistics – Estimates Database (DS-E Database). An innovative database of internationally comparable statistics on disability inequalities. *International Journal of Population Data Science, 8*(6). https://doi.org/10.23889/ijpds.v8i6.2478

Chen, R. J., Wang, J. J., Williamson, D. F. K., Chen, T. Y., Lipkova, J., Lu, M. Y., Sahai, S., & Mahmood, F. (2023). Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nature Biomedical Engineering, 7*(6), 719–742. https://doi.org/10.1038/s41551-023-01056-8

Cibrian, F. L., Chen, Y., Anderson, K., Abrahamsson, C. M., & Motti, V. G. (2025). Limitations in speech recognition for young adults with Down syndrome. *Universal Access in the Information Society, 24*(3), 2295–2313. https://doi.org/10.1007/s10209-025-01197-4

Coffey, K. V., Krahn, G. L., Hanley, J. P., & Neely, J. E. (2026). Identifying implicit bias in LLM-based chat AI toward people with intellectual disabilities. *Disability and Health Journal, 19*(4), Article 102086. https://doi.org/10.1016/j.dhjo.2026.102086

Dalvit, L. (2025). Online suffering and disability in South Africa: Three ethical dilemmas and associated risks. *Journal of Communication Inquiry, 49*(4), 440–460. https://doi.org/10.1177/01968599251338292

Esposito, A., Lanzilotti, R., Matera, M., Piccinno, A., Piro, L., Pucci, E., & Ragone, G. (2026). Conversational AI for digital accessibility: An experimental study involving blind and low vision users. *International Journal of Human–Computer Interaction*. Advance online publication. https://doi.org/10.1080/10447318.2026.2659951

Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., Zuiderveen Borgesius, F., & Biega, A. J. (2025). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Transactions on Intelligent Systems and Technology, 16*(1), 1–54. https://doi.org/10.1145/3696457

Fernández Llorca, D., Frau, P., Parra, I., Izquierdo, R., & Gómez, E. (2024). Attribute annotation and bias evaluation in visual datasets for autonomous driving. *Journal of Big Data, 11*(1), Article 137. https://doi.org/10.1186/s40537-024-00976-9

Foley, A., & Melese, F. (2026). Disabling AI: Power, exclusion, and disability. *British Journal of Sociology of Education, 47*(7), 1416–1437. https://doi.org/10.1080/01425692.2025.2519482

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics, 50*(3), 1097–1179. https://doi.org/10.1162/coli*a*00524

Genga, S. (2025). A review of the adequacy of Kenya's and South Africa's data protection legal frameworks in protecting persons with disabilities from artificial intelligence algorithm discrimination. *African Journal on Privacy & Data Protection, 2*(1), 41–60. https://doi.org/10.29053/ajpdp.v2i1.0003

Hanass-Hancock, J., Kamalakannan, S., Murthy, G. V. S., Palmer, M., Pinilla-Roncancio, M., Rivas Velarde, M., Tetali, S., & Mitra, S. (2023). What cut-off(s) to use with the Washington Group short set of questions? *Disability and Health Journal, 16*(4), Article 101499. https://doi.org/10.1016/j.dhjo.2023.101499

Hasegawa-Johnson, M., Zheng, X., Kim, H., Mendes, C., Dickinson, M., Hege, E., Zwilling, C., Channell, M. M., Mattie, L., Hodges, H., Ramig, L., Bellard, M., Shebanek, M., Sarı, L., Kalgaonkar, K., Frerichs, D., Bigham, J. P., Findlater, L., Lea, C., … MacDonald, B. (2024). Community-supported shared infrastructure in support of speech accessibility. *Journal of Speech, Language, and Hearing Research, 67*(11), 4162–4175. https://doi.org/10.1044/2024_JSLHR-24-00122

Hassard, J., Yildrim, M., Thomson, L., & Blake, H. (2024). Disclosing non-visible disabilities in educational workplaces: A scoping review. *British Medical Bulletin, 150*(1), 23–41. https://doi.org/10.1093/bmb/ldae004

İnce, E. (2026). Algorithmic ableism and the normative body: Disability in the age of technocapitalist AI. *Technology and Disability, 38*(4), 381–389. https://doi.org/10.1177/10554181261433364

Kaboré, S. S., Bayani, A., Epoh Ewane, L. P., dos Anjos, D. S. O., Nikiéma, J. N., & Mac-Seing, M. (2026). Contribution of artificial intelligence research to the development of inclusive applications for people with disabilities: A scoping review (2015–2025). *PLOS Digital Health, 5*(10), Article e0001749. https://doi.org/10.1371/journal.pdig.0001749

Liu, J., Luo, X., & Tang, G. H. (2026). Bridging fairness in machine learning: A legal critique of technical strategies. *Computer Law & Security Review, 61*, Article 106339. https://doi.org/10.1016/j.clsr.2026.106339

Luoga, P., Mahiti, G. R., Nyamhanga, T., Adam, J., Kapinga, O., Herzon, B., Mwaisengela, S. M., Materu, P. A., German, C. J., Mtulo, G. E., Ndeki, D. G., Kapologwe, N. A., & Eliakimu, E. S. (2026). Socioeconomic and national disparities in disability distribution among population in five East African countries: Insights from the demographic and health surveys. *International Journal for Equity in Health, 25*(1), Article 61. https://doi.org/10.1186/s12939-026-02781-7

Matos, J., Van Calster, B., Celi, L. A., Dhiman, P., Gichoya, J. W., Riley, R. D., Russell, C., Khalid, S., & Collins, G. S. (2026). Critical appraisal of fairness metrics for artificial intelligence-based clinical prediction models: A scoping review. *The Lancet Digital Health*. Advance online publication, Article 101001. https://doi.org/10.1016/j.landig.2026.101001

Mattioli, M., & Cabitza, F. (2024). Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology. *Machine Learning and Knowledge Extraction, 6*(4), 2201–2231. https://doi.org/10.3390/make6040109

McInerney, K., & Keyes, O. (2025). The infopolitics of feeling: How race and disability are configured in emotion recognition technology. *New Media & Society, 27*(7), 4246–4264. https://doi.org/10.1177/14614448241235914

Mirsch, M., Wegner, L., Strube, J., & Leicht-Scholten, C. (2026). Toward substantive intersectional algorithmic fairness: Desiderata for a feminist approach. *AI and Ethics, 6*(1), Article 90. https://doi.org/10.1007/s43681-025-00926-y

Mladenov, T. (2026). AI and disabled people's independent living: A framework for analysis. *AI & Society, 41*(3), 2593–2605. https://doi.org/10.1007/s00146-025-02642-x

Moura, I. (2023). Encoding normative ethics: On algorithmic bias and disability. *First Monday, 28*(1). https://doi.org/10.5210/fm.v28i1.12905

Moura, I., Powell, R. M., & Kaye, H. S. (2026). Evaluating disability bias in the Allegheny Family Screening Tool: Implications for child welfare decisions. *Journal of Public Child Welfare*. Advance online publication. https://doi.org/10.1080/15548732.2026.2689958

Mutia, F., Aysya, A. A. A., Fahmi, F., & Atmi, R. T. (2026). Uncovering ableism in large language models' responses: Hybrid sentiment and thematic analysis approach for disability representation. *Social Sciences, 15*(7), Article 470. https://doi.org/10.3390/socsci15070470

Newman-Griffis, D., Rauchberg, J. S., Alharbi, R., Hickman, L., & Hochheiser, H. (2023). Definition drives design: Disability models and mechanisms of bias in AI technologies. *First Monday, 28*(1). https://doi.org/10.5210/fm.v28i1.12903

Pierrès, O., Darvishy, A., & Christen, M. (2025). Exploring the role of generative AI in higher education: Semi-structured interviews with students with disabilities. *Education and Information Technologies, 30*(7), 8923–8952. https://doi.org/10.1007/s10639-024-13134-8

Plantinga, P., Shilongo, K., Mudongo, O., Umubyeyi, A., Gastrow, M., & Razzano, G. (2024). Responsible artificial intelligence in Africa: Towards policy learning. *Data & Policy, 6*, Article e72. https://doi.org/10.1017/dap.2024.60

Pucci, E., Piro, L., Possaghi, I., Mulfari, D., & Matera, M. (2024). Co-designing the integration of voice-based conversational AI and web augmentation to amplify web inclusivity. *Scientific Reports, 14*(1), Article 16162. https://doi.org/10.1038/s41598-024-66725-3

Rigotti, C., & Fosch-Villaronga, E. (2024). Fairness, AI & recruitment. *Computer Law & Security Review, 53*, Article 105966. https://doi.org/10.1016/j.clsr.2024.105966

Santuzzi, A. M., Keating, R. T., & Martinez, J. J. (2025). The impact of response options on formal disclosure rates for disabilities in organizations. *Group & Organization Management, 50*(4), 1097–1110. https://doi.org/10.1177/10596011241258669

Selwyn, N., O'Neill, C., Smith, G., Andrejevic, M., & Gu, X. (2023). A necessary evil? The rise of online exam proctoring in Australian universities. *Media International Australia, 186*(1), 149–164. https://doi.org/10.1177/1329878X211005862

Sohrabi, R. (2026). When AI fails to listen: Convergent failure patterns across speech impairments and low-resource speech in ASR. *AI & Society, 41*(9), 8811–8821. https://doi.org/10.1007/s00146-026-03073-y

Umucu, E. (2025). Artificial intelligence and health equity for people with disabilities: An integrated framework for disability-inclusive AI design. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 62*, Article 00469580251365472. https://doi.org/10.1177/00469580251365472

Zhong, Z., Wang, Q., Singh, S., Mendes, C. C., Hasegawa-Johnson, M., Abdulla, W., & Shahamiri, S. R. (2025). Convolution-augmented transformers for enhanced speaker-independent dysarthric speech recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 33*, 3815–3826. https://doi.org/10.1109/TNSRE.2025.3610792