A Systematic Literature Review on Bias Evaluation and Mitigation in Automatic Speech Recognition Models for Low-Resource African Languages

JOYCE NAKATUMBA-NABENDE, Makerere University, Kampala, Uganda and Makerere University Center for Artificial Intelligence, Kampala, Uganda
SULAIMAN KAGUMIRE, Makerere University CEDAT, Kampala, Uganda
CAROLINE KANTONO, Center for Artificial Intelligence and Data Science, Makerere University, Kampala, Uganda
PETER NABENDE, Information Systems, Makerere University, Kampala, Uganda

With recent advancements in speech recognition, it is crucial to ensure that automatic speech recognition (ASR) systems do not exhibit systematic biases, such as those related to gender, age, accent, and dialect. Although research has extensively examined systematic biases such as those related to gender, age, accent, and dialect, for high-resource languages, research on low-resource African languages remains limited. This systematic literature review synthesizes evidence on bias evaluation and mitigation in ASR models for African languages, adhering to the PRISMA reporting guidelines. Our analysis reveals that most biases stem from data imbalances and limited linguistic diversity in training datasets, resulting in disproportionately high error rates for underrepresented speaker groups. Mitigation strategies in African contexts have primarily focused on data-centric methods, including dataset expansion, augmentation, and transfer learning. In contrast, more advanced approaches, including fairness-aware modeling, bias-aware loss functions, adversarial debiasing, and speaker-adaptive techniques, are rarely applied. Gender, accent, and dialect biases dominate the few African studies available, while age and racial biases are almost absent. The limited number of African languages covered highlights the urgent need for more representative and inclusive research. Addressing these gaps will support the development of fairer and more robust ASR technologies across the continent.

CCS Concepts: • **Computing methodologies → Machine learning approaches**; **Speech recognition**.

Additional Key Words and Phrases: Automatic Speech Recognition, Bias, Fairness, Low-Resource African Languages, Bias Evaluation, Bias Mitigation

1 Introduction

Automatic Speech Recognition (ASR) has reshaped how we interact with technology, powering tools such as virtual assistants and transcription tools. Despite their ubiquity, ASR systems still have limitations. A major limitation of ASR systems is the presence of biases in training data and model design, which can degrade their performance. For instance, research has reported that ASR systems demonstrate gender bias, leading to higher error rates for female speakers [22, 30]. Beyond gender, ASR systems are also significantly biased by a speaker's dialect, accent, age, and race, which degrade the speaker's performance for groups such as those with non-standard

Authors' Contact Information: Joyce Nakatumba-Nabende, Makerere University, Kampala, Uganda and Makerere University Center for Artificial Intelligence, Kampala, Uganda; e-mail: joyce.nabende@mak.ac.ug; Sulaiman Kagumire, Makerere University CEDAT, Kampala, Central Region, Uganda; e-mail: sulaiman.kagumire@gmail.com; Caroline Kantono, Center for Artificial Intelligence and Data Science, Makerere University, Kampala, Central Region, Uganda; e-mail: ckantono100@gmail.com; Peter Nabende, Information Systems, Makerere University, Kampala, Central Region, Uganda; e-mail: peter.nabende@mak.ac.ug.

ACM 1557-7341/2025/9-ART
https://doi.org/10.1145/3769089

dialects or with unique accents [67]. Research has been conducted to understand these disparities, particularly for high-resource languages. Several strategies to mitigate these biases and improve the overall accuracy and inclusivity of ASR systems have been developed [1, 19, 24, 28, 42, 86].

Unfortunately, progress in mitigating biases in Automatic Speech Recognition (ASR) has not extended to low-resource languages, particularly those spoken in Africa. This challenge arises due to a lack of representative datasets, which are often skewed toward single genders or standard accents, thereby marginalizing many speakers. To address this critical gap, this systematic literature review (SLR) maps the current landscape of understanding biases across gender, dialect, age, and accent in ASR models for African languages. This literature review synthesizes existing work to understand ASR biases in low-resource African languages. The existing disparity risks exacerbating existing social inequalities by excluding marginalized groups, including children, older speakers, and those with diverse dialects or accents, from the benefits provided by ASR systems [2, 64]. As technology increasingly shapes access to information and opportunities, building fair and inclusive ASR systems is not just a technical challenge; it is also a social imperative [57]. This literature review presents an overview of the current state of ASR bias research for low-resource African languages, identifying critical research gaps and promising mitigation strategies. This research aims to guide future studies that prioritize equity, ensuring that technological advancements benefit everyone, regardless of their gender, dialect, age, or accent. Our research aims to answer the following questions:

- *RQ1: What types of biases have been identified in ASR models for low-resource African languages?*
- *RQ2: What strategies have been proposed to mitigate these biases and promote inclusion in ASR systems?*

The results of the review show five categories of ASR bias: gender, accent, dialect, age, and racial/ethnic bias. The research focus, however, differs dramatically between linguistic contexts. In high-resource languages, studies have concentrated on gender and accent, consistently finding higher error rates for women and speakers of non-standard accents. In contrast, research on African languages is significantly smaller and narrower. The few existing studies have mainly addressed accent and dialect bias, with less attention to gender, while age-related and racial/ethnic biases remain almost entirely unexplored. The paper's findings highlight data-centric approaches as the primary method for addressing accent and dialect biases in ASR models. Model-centric techniques, such as finetuning and transfer learning, were also observed, often targeting gender and accent disparities. There were no mitigation strategies in place to address age or racial biases. These findings demonstrate that while current research strongly emphasizes fairness for accents, dialects, and gender, other types of bias remain largely unstudied.

The rest of the paper is organized as follows. Section 2 presents the related work section. Section 3 provides the methodology used in the paper. Sections 4 and 5 highlight the main results from the systematic review and the discussion of the results, respectively. Finally, Section 6 provides a conclusion to the paper.

2 Related Work

Numerous studies in Natural Language Processing (NLP) have extensively examined biases across gender, age, dialect, and accent. Existing research primarily addresses these biases within tasks such as machine translation [3, 35, 66, 70, 72, 77, 89] and sentiment analysis [4, 49]. The development of Automatic Speech Recognition (ASR) models for African languages is gaining momentum with varying regional focus. East African efforts have concentrated on languages like Luganda, Kinyarwanda, and Swahili [7, 13, 17, 41, 55, 61, 64, 65]. Similarly, West African ASR development includes languages such as Fon, Igbo, Yoruba, Wolof, Bambara, Maasina Fulfulde, Asante-Twi, and Nigerian Pidgin [16, 26, 29, 58]. Southern African research has focused on Zambian languages (Bemba, Nyanja, Tonga, Lozi, Lunda) [79] and South African languages (isiZulu, isiXhosa, Sesotho, Setswana) [95]. Recent studies have also explored ASR for Congolese languages (Lingala, Kongo/Kikongo, Congolese Swahili, LubaKasai/Tshiluba) [39] and Ethiopian languages (Amharic, Oromo, Tigrinya, Wolaytta, Somali) [58, 84]. This growing body of work shows the increasing attention to ASR development in diverse African linguistic contexts.

2.1 Biases in Automatic Speech Recognition Systems

Automatic Speech Recognition (ASR) systems exhibit bias when their performance varies systematically across user demographics [96], typically evidenced by unequal Word Error Rates (WER) among different speaker groups [20]. Research has identified several prominent forms of biases in ASR models. Gender bias refers to disparities in recognition accuracy between male and female speakers [19, 78].

Gender bias in ASR technology affects both the technical performance of the models and social interactions among humans. On a technical level, commercial systems from major providers such as Google and Amazon consistently have higher error rates for women compared to men [67]. Research has confirmed, for instance, that voice assistants such as Siri systematically underperform for female voices [5]. On a social level, voice assistants like Alexa and Siri often reinforce harmful stereotypes. This underperformance can occur through design choices, such as defaulting to female voices in service-oriented roles, and through alarming interactional failures, like responding inappropriately to harassment [10, 45, 82]. Together, these failures in performance and design demonstrate how ASR technologies not only replicate societal biases but also actively amplify them.

Accent bias arises when ASR models underperform for accented speech due to under-representation in training data, leading to systematically higher error rates for speakers of non-standard or regional accents [20, 98]. *Age bias* stems from variations in vocal tract length, pitch, and articulation across age groups. Models trained predominantly on adult speech often fail to capture the acoustic characteristics of children, producing markedly higher error rates for younger speakers [9, 92].

Dialect bias occurs when ASR models favor mainstream urban dialects over regional or rural ones due to skewed data collection. In the UK, for example, ASR systems struggle with regional dialects, such as those found in Newcastle and Scottish English, compared to the standard accent [76, 87]. Data from the United States also highlights this disparity, showing that speakers of African American and Yakama dialects face much higher error rates than White speakers [90]. These biases stem directly from urban-centric data practices that over-represent city speech while neglecting rural and minority varieties. Dialect bias is also relevant in the African context, where, for example, training data for ASR models often prioritizes standardized forms, such as urban Kiswahili, thereby effectively excluding remote dialects.

Racial bias reflects systematic inaccuracies in recognising speech from a specific racial or ethnic group. Racial bias in ASR systems creates significant performance gaps, with studies showing commercial models produce nearly twice as many errors for African American speakers as for White speakers [42, 51]. Solving racial bias requires more than just technical fixes; it demands a fundamental shift in how data is collected and analysed. For example, research on gathering African American Vernacular English (AAVE) data highlights the need for community-centred methods that build trust and ensure equitable representation [37]. Therefore, overcoming racial bias in ASR requires a combination of technical improvements and inclusive, participatory approaches to dataset creation. Overall, extensive research has been conducted to study demographic and linguistic biases in high-resource languages; however, low-resource African languages remain largely underexplored. The research trend is particularly concerning for African languages, where limited datasets and urban-centric data collection practices risk amplifying existing inequalities rather than enabling inclusive speech technologies.

2.2 Mitigation Strategies for Bias in ASR and NLP Models

Beyond identifying and evaluating biases, numerous studies have explored mitigation techniques in ASR and NLP. Systematic reviews highlight various approaches, including data augmentation, adversarial training, and embedding debiasing [83]. Notably, many of these methods have been primarily tested on English, raising questions about their efficacy in low-resource African languages. Reviews targeting ASR biases, such as gender, racial, and disability-related disparities, examine strategies like data augmentation, improved datasets, adversarial training, and fairness-aware model adaptation [67].

There is a need for inclusive datasets and accessibility for marginalized communities [92]. Techniques such as vocal tract length normalization (VTLN) and model adaptation are discussed in the context of children's speech recognition [50]. Analyses of predictive bias in English ASR highlight the crucial role of data bias and advocate for proactive language management and the inclusion of diverse datasets. Studies focusing on African American Language (AAL) highlight systematic performance disparities and call for linguistically inclusive datasets that integrate sociolinguistic research into model development. Mitigation strategies such as data augmentation with diverse linguistic features and bias-aware model training have been proposed to address racial disparities in speech recognition systems.

While existing literature extensively explores biases in ASR systems, a critical gap remains: the lack of systematic reviews focusing on low-resource languages, particularly those in Africa. Furthermore, researchers have extensively analyzed gender, accent, age, and racial biases in languages like English, however, there is limited research on bias mitigation for African languages. This study addresses this gap by conducting the first systematic review specifically investigating biases in ASR models for low-resource African languages. It aims to assess the applicability of current mitigation strategies and provide a foundation for future research on fairness and inclusivity in these underrepresented, low-resourced African languages.

3 Research Methodology

This systematic literature review adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol [59], ensuring a transparent and replicable selection process. The methodology comprised four key steps: formulating clear research questions, developing a comprehensive search strategy, establishing explicit inclusion and exclusion criteria, and conducting a rigorous quality assessment followed by data extraction.

3.1 Research Questions (RQs)

This systematic literature review (SLR) aims to comprehensively investigate biases within ASR systems, with a specific focus on low-resource African languages. We examined key types of bias reported in the literature, including gender, accent, dialect, age, and racial/ethnic bias. We analyze how these biases manifest themselves in ASR models, evaluate their impact on system performance, and assess mitigation strategies proposed in previous work. To guide this investigation, we formulate the following research questions (RQs):

- *RQ1: What types of biases have been identified in ASR models for low-resource African languages?*
- *RQ2: What strategies have been proposed to mitigate these biases and promote inclusion in ASR systems?*

3.2 Search Strategy

We conducted an extensive search across multiple academic databases to achieve a comprehensive and systematic retrieval of relevant papers. Our search strategy entailed 1) selecting relevant academic databases, 2) formulating search queries using relevant keywords and Boolean operators, 3) applying appropriate search filters to refine results, and 4) systematically documenting all retrieved studies.

3.2.1 Electronic Databases. We selected five well-known databases widely utilized in Natural Language Processing (NLP) and Machine Learning research to ensure a robust search for relevant literature. This study focused on peer-reviewed journal articles, conference proceedings, and workshop papers published between 2000 and 2025. We chose the databases for their diverse and extensive collections of high-quality, peer-reviewed publications, including systematic literature reviews: (1) IEEE Xplore digital library(http://ieeexplore.ieee.org),

(2) ACM digital library (https://dl.acm.org), (3) ScienceDirect (http://www.sciencedirect.com), (4) SpringerLink (https://link.springer.com) and Google Scholar (https://scholar.google.com). We employed the search keywords detailed in Table 1 and retrieved 2,829 articles.

3.2.2 Search Query Construction. We formulated the search queries using Boolean operators (AND, OR) to refine the search process. The keywords depicted in Table 1 were selected to cover various forms of biases in ASR: gender, accent bias, age, and dialect.

Table 1. Search Keywords and Query Strings for Systematic Review.

Keywords Group	Search String
Bias in ASR	"Gender Bias" OR "Accent Bias" OR "Dialect Bias" OR "Age Bias" in "ASR"
Bias	"ASR for low resource languages"
Inclusivity in ASR	"Fairness" AND "Inclusivity" in "ASR"
Mitigating Bias in ASR	"Mitigating Bias" in "ASR" OR "Speech Recognition"

3.3 Article Selection Process

Our article selection process consisted of five steps, as shown in Figure 1. The figure details the number of papers retrieved, screened, and ultimately included in the final synthesis. Initially, we selected a total of 2,829 papers.

3.3.1 Inclusion and Exclusion Criteria. To ensure a rigorous selection of studies, we applied inclusion and exclusion criteria, which enabled us to filter out and obtain high-quality, peer-reviewed papers directly relevant to our research questions. Table 2 provides a comprehensive overview of the criteria. In this step, we removed 1,962 papers and selected 867 articles for abstract screening.

Table 2. Inclusion and Exclusion Criteria for Study Selection.

Inclusion Criteria	Exclusion Criteria
Journal articles, conference papers, and workshop papers	Books, theses, dissertations, blogs, prefaces
The paper is accessible with the full text available	Full text is inaccessible
Papers must be written in English	Non-English papers
Papers published between 2000 and 2025	Duplicates

3.3.2 Abstract Screening. The 867 papers that passed the initial identification phase were screened based on the relevance of their abstracts to the aims of this research. Each abstract was reviewed at this stage to determine its alignment with the research questions, with a primary focus on studies addressing biases in ASR models. During this phase, 681 papers were excluded based on the following criteria:

- *Keywords not relevant* (n = 151): The abstract did not contain any reference to bias in ASR models, fairness, or inclusivity.
- *Papers out of scope* (n = 427): The study focused on non-related NLP tasks (e.g., machine translation, sentiment analysis, facial recognition).
- *Insufficient information in the abstract* (n = 103): The abstract did not provide clear details on the study's research objectives, methodology, or findings related to ASR bias.

After applying the exclusion criteria, 186 papers remained. We then reviewed the full text of these articles to assess them against our inclusion criteria.

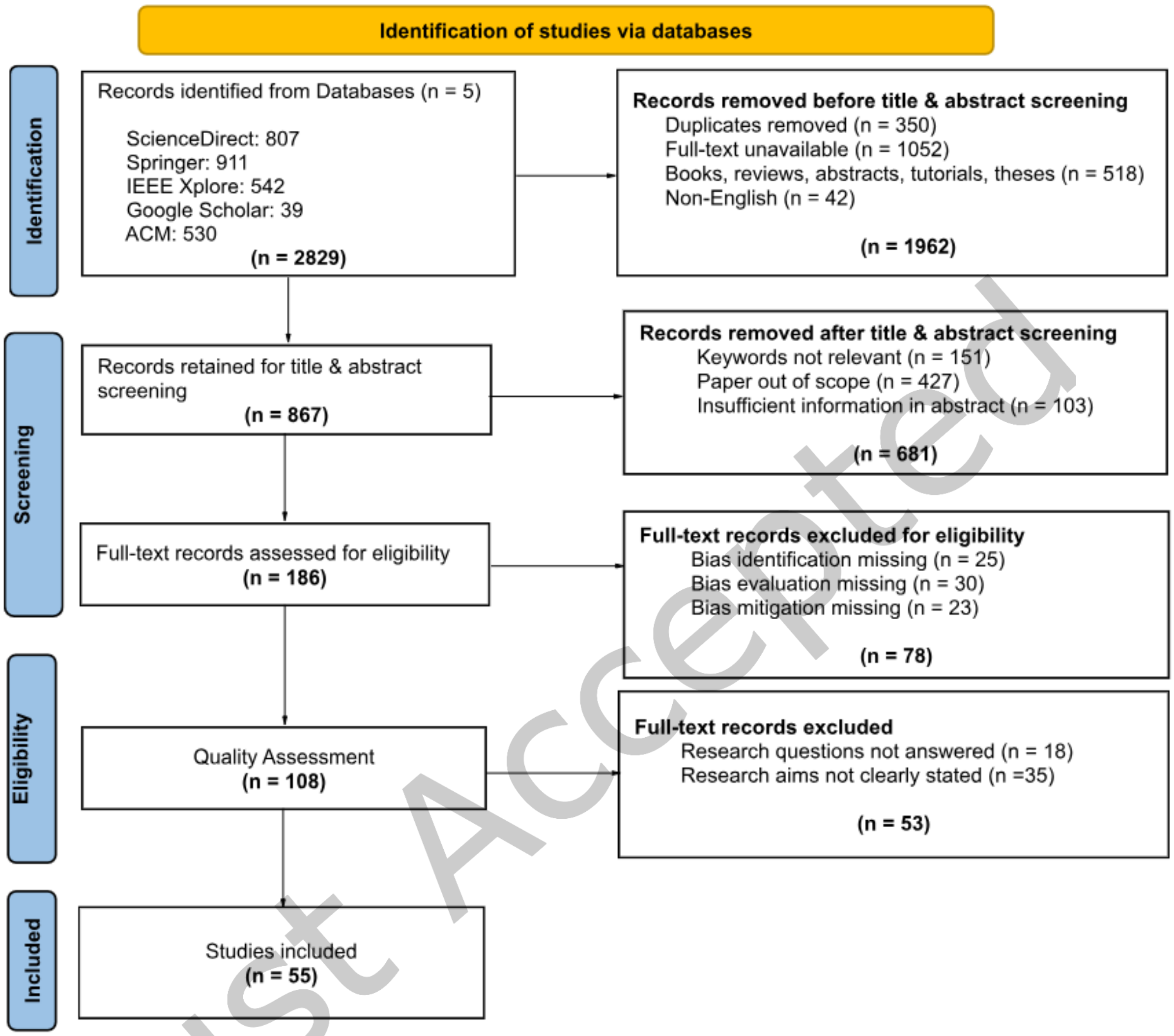

Fig. 1. PRISMA flow chart of the article selection process.

3.3.3 Full-text Screening. At this stage, we carefully reviewed 186 papers to determine if they provided substantive information on the identification, evaluation, or mitigation of ASR bias. We retained 108 papers and excluded 78 papers for the following reasons:

- *Bias identification missing* (n = 25): The study did not explicitly discuss or analyze bias in ASR models.
- *Bias evaluation missing* (n = 30): Although the study mentioned bias, it lacked quantitative or qualitative assessment metrics for the identified bias.
- *There were no bias mitigation strategies proposed* (n = 23): The paper focused only on the existence of bias without suggesting mitigation techniques.

3.4 Quality Assessment

To ensure methodological accuracy and relevance, we performed a quality assessment of each paper. Following the systematic review guidelines from [40], we evaluated each paper against predefined criteria, including only methodologically sound papers in our final evaluation and analysis.

- *Q1:* Does the paper directly address bias identification, evaluation, or mitigation in ASR models?
- *Q2:* Is the paper design suitable for answering the research questions?
- *Q3:* Does the paper clearly describe its research objectives and methodology?
- *Q4:* Does the paper clearly describe the data sources?
- *Q5:* Are fairness metrics, such as WER disparities, speaker demographics, or fairness-aware training techniques, applied?
- *Q6:* Does the paper document the methods, datasets, and models?
- *Q7:* Does the paper discuss its limitations and possible sources of bias?

We excluded any paper that failed to meet at least three of the seven quality criteria. All exclusion decisions were made by consensus between two reviewers to ensure consistency and transparency. As a result of this quality assessment, 53 papers were removed, most commonly because:

- Research questions were not explicitly answered (n = 18).
- Research aims were not clearly stated (n = 35).

After applying the inclusion, exclusion, and quality assessment criteria, we selected a final set of 55 papers for the systematic review.

3.5 Data Extraction

After selecting the 55 studies, we extracted and synthesized the data to answer our research questions (Section 3.1). We analyzed each study to identify its context (low-resource vs. high-resource languages) and systematically coded it for its focus on gender, age, accent, racial, or dialect bias. Furthermore, each article was coded for its inclusion of bias evaluation and mitigation techniques.

Our analysis revealed several key findings, as detailed in Table 3. All 55 articles addressed at least one of the five bias types, and every study proposed at least one bias mitigation approach, showing a strong research emphasis on enhancing ASR fairness. While most studies included a biased evaluation, one article focused solely on proposing a mitigation strategy without explicit evaluation. This variation highlights the diverse approaches in the field, with some studies prioritizing the identification of bias and others focusing on practical reduction strategies. Figure 2 presents the research papers by publication type (Journal, Conference, Workshop) from 2000 to 2025. In general, conferences represented the majority of publications (51.7%), followed by journals (34.5%) and workshops (13.8%).

4 Results

This section discusses the results obtained for each research question mentioned in Section 3.1.

4.1 Biases and Inclusivity in ASR

This section aims to answer our first research question.
RQ1: What are (gender) biases and inclusivity of Automatic Speech Recognition models developed for low-resource languages on the African continent? To address this question, we investigated how different biases have been studied in ASR research. Specifically, we examined the number of papers focusing on each bias type and whether they target high-resource languages or low-resource African languages.

Table 3. Distribution of papers investigated and reviewed in the systematic literature review.

Ref	Language	Gender Bias	Age Bias	Accent Bias	Racial Bias	Dialect Bias	Bias Evaluation	Bias Mitigation
[54]	English				✓		✓	✓
[62]	Korean	✓				✓	✓	✓
[93]	English					✓	✓	✓
[73]	English			✓			✓	✓
[61]	Luganda	✓					✓	✓
[71]	English			✓			✓	✓
[14].	English			✓			✓	✓
[69]	English			✓			✓	✓
[56]	English	✓	✓	✓			✓	✓
[68]	Nigerian-Accent English			✓			✓	✓
[8]	Nigerian-Accented English			✓			✓	✓
[52]	English				✓		✓	✓
[33]	Flemish	✓	✓	✓		✓	✓	✓
[38]	English			✓			✓	✓
[99]	Dutch			✓			✓	✓
[80]	Kiswahili					✓	✓	✓
[74]	English	✓	✓		✓	✓	✓	✓
[15]	English					✓	✓	✓
[1]	English, French	✓					✓	✓
[53]	English	✓					✓	✓
[86]	English				✓	✓	✓	✓
[28]	English	✓		✓			✓	✓
[34]	English			✓			✓	✓
[96]	Flemish Dutch			✓			✓	✓
[100]	French, Italian, Baque, Portuguese, Catalan							✓
[11]	Mexican Spanish	✓					✓	✓
[88]	Finnish	✓	✓				✓	✓
[63]	Spanish	✓		✓			✓	✓
[85]	English	✓				✓	✓	✓
[31]	Bangla	✓					✓	✓
[24]	French	✓					✓	✓
[101]	English	✓					✓	✓
[94]	English			✓			✓	✓
[25]	English	✓					✓	✓
[91]	English			✓			✓	✓
[81]	English		✓				✓	✓
[97]	Dutch			✓			✓	✓
[30]	English	✓				✓	✓	✓
[6]	19 languages	✓					✓	✓
[21]	English	✓					✓	✓
[42]	English				✓		✓	✓
[23]	Xhosa, other multi-dialect English					✓	✓	✓
[60]	Moroccan Arabic					✓	✓	✓
[32]	Ghanaian English			✓			✓	✓
[43]	Portuguese	✓	✓				✓	✓
[75]	Arabic	✓	✓	✓			✓	✓
[47]	Irish					✓	✓	✓
[19]	Dutch, Mandarin	✓	✓	✓		✓	✓	✓
[44]	English	✓					✓	✓
[22]	Dutch	✓	✓	✓			✓	✓
[27]	English	✓					✓	✓
[36]	English	✓	✓	✓			✓	✓
[48]	Irish					✓	✓	✓
[12]	English			✓			✓	
[18]	Brazilian Portuguese	✓	✓	✓			✓	✓

Figure 3 provides an overview of the distribution of bias studies in ASR. The left subplot illustrates the bias distribution in high-resource languages (e.g., English, French, Spanish), while the right subplot highlights studies focusing on low-resource African languages. Our study reveals that **accent, gender, age, dialect, and racial** biases are the most studied biases in ASR research. However, there is a clear contrast in their distribution between high-resource and low-resource African languages. Among the studies on high-resource languages, accent and

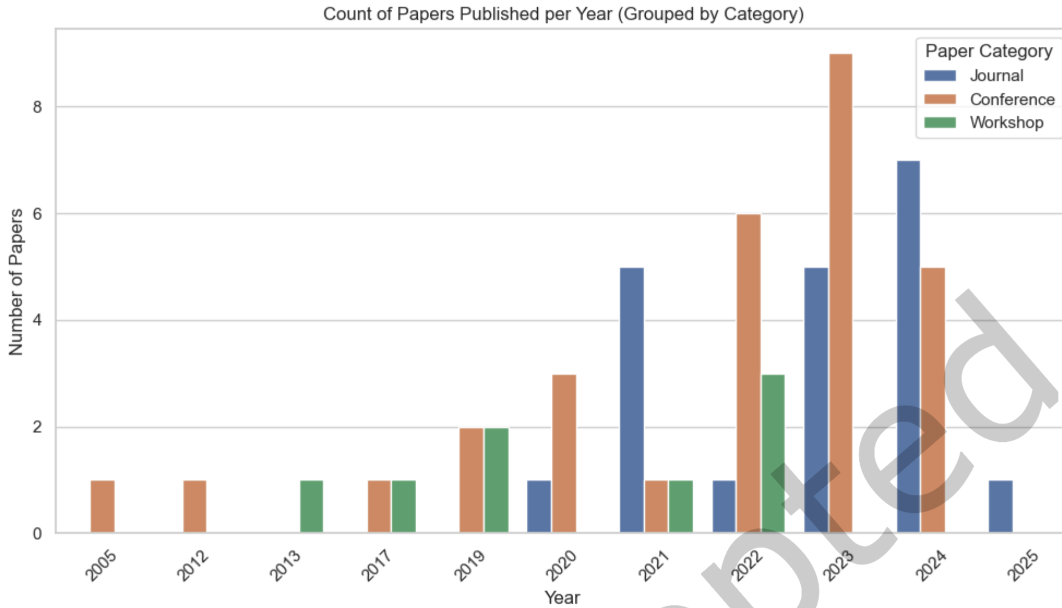

Fig. 2. Distribution of the research papers highlighting research for African languages.

gender biases were the most frequently investigated, appearing in 23 and 26 papers, respectively. Age and dialect biases were explored in 11 and 12 papers, respectively, while racial bias was the least studied, with only five papers.

Our review found that research on ASR bias in low-resource African languages is limited. Of the seven studies we identified, the primary focus was on accent and dialect bias (three papers each), with only one study addressing gender bias. Critically, we found no research that examined age or racial biases. This absence highlights a significant gap in fairness-aware development of equitable ASR for underrepresented linguistic communities.

4.1.1 Accent Bias: Several studies highlight the issue of accent bias in ASR systems, particularly within African linguistic contexts. The authors in [68] demonstrated how ASR models trained on Nigerian-accented English struggled to differentiate between speakers from various ethnic groups, revealing inherent biases in model design. Their findings showed that Gaussian Mixture Models (GMMs) frequently misclassified speakers, while Logistic Regression, though more accurate, still faced challenges distinguishing between accents. Babatunde et al. [8] further explored this issue using transfer learning techniques, showing that while models like QuartzNet achieved relatively low Word Error Rates (WERs), commercial ASR APIs performed poorly on Nigerian-accented speech. Henkel et al. [32] investigated ASR models for assessing oral reading fluency in Ghanaian students, showing that Whisper V2 outperformed Wav2Vec2.0, indicating that modern models can achieve reasonable WER in Ghana English. These studies highlight the challenge of developing ASR systems that accurately recognise diverse accents within a single country.

4.1.2 Dialect Bias: Dialect bias presents another significant challenge in developing inclusive ASR systems. Siminyu et al. [80] highlighted the potential for dialect bias when creating an ASR corpus for Kiswahili, noting that models trained on Standard Kiswahili may fail to generalize across regional dialects. Gao et al. [23] identified accuracy disparities across English dialects, including Xhosa English, and introduced the Equal Accuracy Ratio

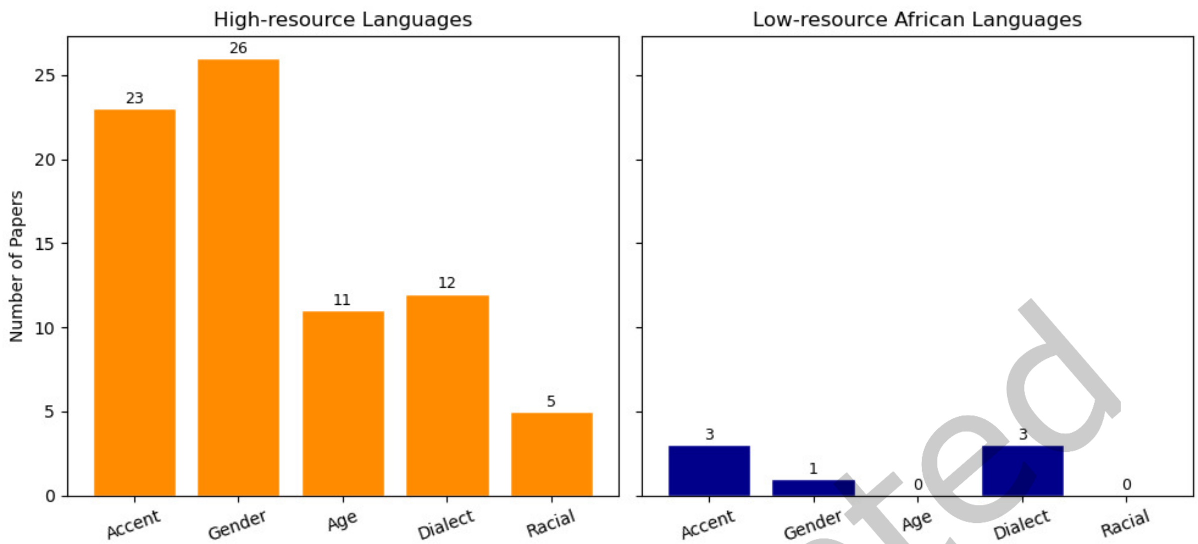

Fig. 3. Bias study distribution in high-resource languages (left) and low-resource African languages (right). The y-axis represents the number of papers addressing each bias.

(EAR) to address these disparities. Their findings emphasized that dialects with limited training data exhibited higher WERs, and disparities persisted despite joint training. Mouaz et al. [60] reinforced these challenges in dialects by demonstrating the performance gap between Modern Standard Arabic (MSA) and Moroccan dialectal speech, showing that MSA-trained models failed to recognize dialect Arabic speech. These studies collectively illustrate the challenges in developing ASR systems that are both fair and accurate across diverse dialects.

4.1.3 Gender Bias: Gender bias remains a critical concern in ASR development, particularly in low-resource settings. Mukiibi et al. [61] developed a Luganda radio speech corpus and found significant gender bias, with female speakers exhibiting higher WERs due to under-representation in the training data. The lack of female voices in the training data set created an additional layer of gender bias in the study in [68]. Siminyu et al. [80] also noted significant biases in ASR performance, aligning with previous findings that speech models tend to be more accurate for men and younger speakers. These studies show the importance of balanced datasets and the need for deliberate efforts to mitigate gender bias in ASR systems.

RQ1: *What types of biases have been identified in ASR models for low-resource African languages?*

Summary for RQ1: *Research on ASR bias has primarily identified five key categories of bias: gender, accent, dialect, age, and racial/ethnic bias. In high-resource languages, studies have concentrated on gender and accent, consistently finding higher error rates for women and speakers of non-standard accents. In contrast, research on African languages is significantly smaller and more focused. The few existing studies have mainly addressed accent and dialect bias, with less attention to gender, while age-related and racial/ethnic biases remain almost entirely unexplored.*

4.2 Bias Mitigation Strategies in ASR Models

To answer *RQ2: What strategies have been proposed to mitigate these biases and promote inclusion in ASR systems?*, we systematically analyzed the mitigation techniques reported in the 55 reviewed studies. Table 4 provides a structured overview of these strategies, grouped into consistent themes that align with the sub-sections below. This ensures that each mitigation theme in the text is directly reflected in the table.

Table 4. Comparison of different ASR bias mitigation strategies, highlighting a brief description of each strategy and reference papers that discuss each of the strategies.

Bias Mitigation Strategy	**Brief Description**	**Reference Papers**
Data Diversification	Expanding and diversifying training and evaluation datasets, creating socio-linguistically stratified corpora, and developing multi-dialect and gender-balanced datasets.	[18, 23, 24, 48, 68, 85]
Data Augmentation	Accent, speed, pitch, volume, and spectral perturbations; synthetic speech augmentation; gender-adaptive augmentation.	[8, 24, 27, 96, 100]
Data Balancing	Oversampling techniques, phonetic balancing, and dialect-weighted corpus adjustments.	[43, 47, 56, 75]
Linguistic Data Refinement	Inclusive dataset construction with linguistic engagement; phoneme-level analysis of underrepresented speech.	[22, 52]
Model Finetuning & Adaptation	Finetuning for accents, dialects, gender, domain-specific speech; fairness-aware training; speaker adaptation techniques.	[8, 19, 25, 32, 61]
Transfer Learning	Cross-lingual adaptation and multilingual pretraining transfer; adaptation from high-resource to low-resource languages.	[27, 96, 99]
Model Architecture & Training	Dialect-aware embeddings; accent feature extraction; zero-shot adaptation; self-supervised and meta-learning approaches.	[28, 71, 93, 94, 101]
Advanced Model Techniques	Adversarial training; counterfactual augmentation; intermediate CTC optimization, multi-task, and few-shot learning.	[46, 74, 81, 91]
Fairness-Aware Approaches	Developing gender- or bias-aware metrics; bias-aware model adaptation; incorporating voice timbre and fairness constraints.	[6, 11, 24, 31]
Evaluation & Refinement	Alternative evaluation metrics, domain- and language-specific benchmarking, model optimization via complexity reduction and normalization.	[11, 31, 34]

Our synthesis identified nine broad categories of strategies: (i) Data Diversification, (ii) Data Augmentation, (iii) Data Balancing, (iv) Linguistic Data Refinement, (v) Model Finetuning and Adaptation, (vi) Transfer Learning,

(vii) Model Architecture and Training, (viii) Advanced Model Techniques, (ix) Fairness-Aware Approaches, and (x) Evaluation and Refinement.

Figures 4 and 5 visualize the prevalence of these strategies in low-resource versus high-resource African ASR research. In general, our findings show that data-centric strategies (diversification, increase, and balancing) are the most common, reflecting the critical role of data set representation in mitigating bias. Model-centric techniques, such as finetuning, transfer learning, and architecture modifications, are also widely explored, particularly in African contexts, where researchers adapt pre-trained multilingual models. In contrast, fairness-aware training, advanced techniques, and new evaluation metrics remain underutilized in African ASR, despite growing adoption in high-resource language studies.

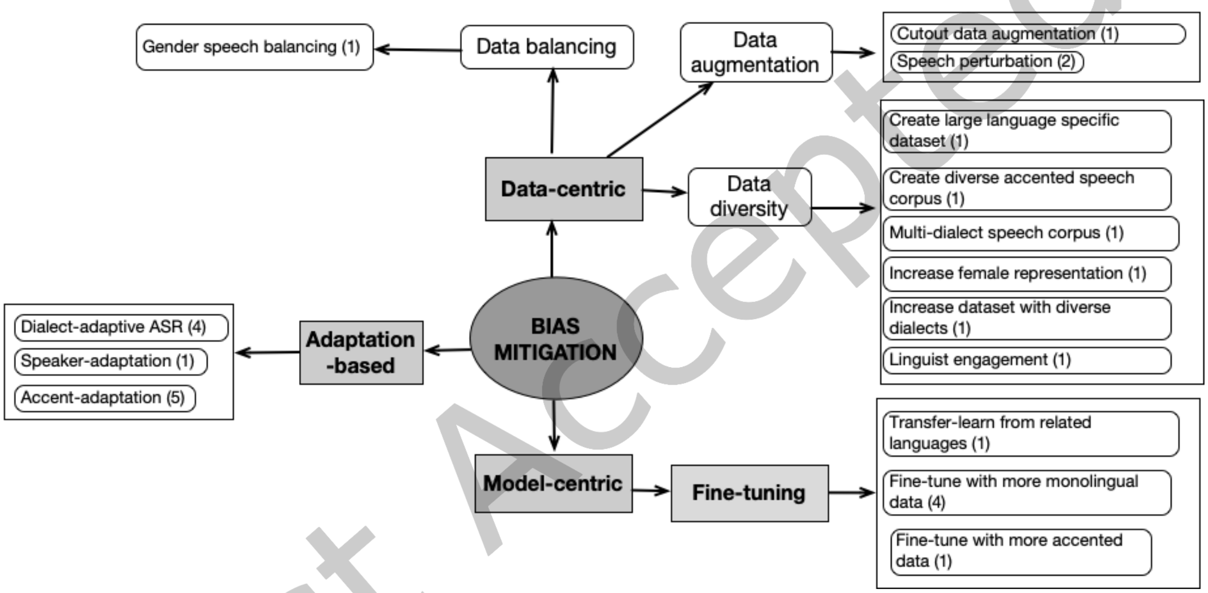

Fig. 4. Distribution of bias mitigation strategies based on research papers on low-resourced African languages. The strategies fall into three main categories: data-centric, model-centric, and adaptation-based.

Our findings reveal that the most common mitigation strategies suggested in high-resource and low-resource African ASR research are data-centric techniques, particularly data augmentation and dataset diversity. These approaches reflect a shared recognition across both research streams that biases often originate from unbalanced or under-representative training data. However, high-resource studies demonstrate broader methodological diversity. For instance, eight studies proposed developing new fairness-aware evaluation metrics beyond traditional WER and CER, emphasizing the limitations of standard metrics in capturing bias-related disparities. In contrast, low-resource African studies relied more heavily on transfer learning and finetuning, which indicates limited access to large-scale, diverse language corpora and computational resources.

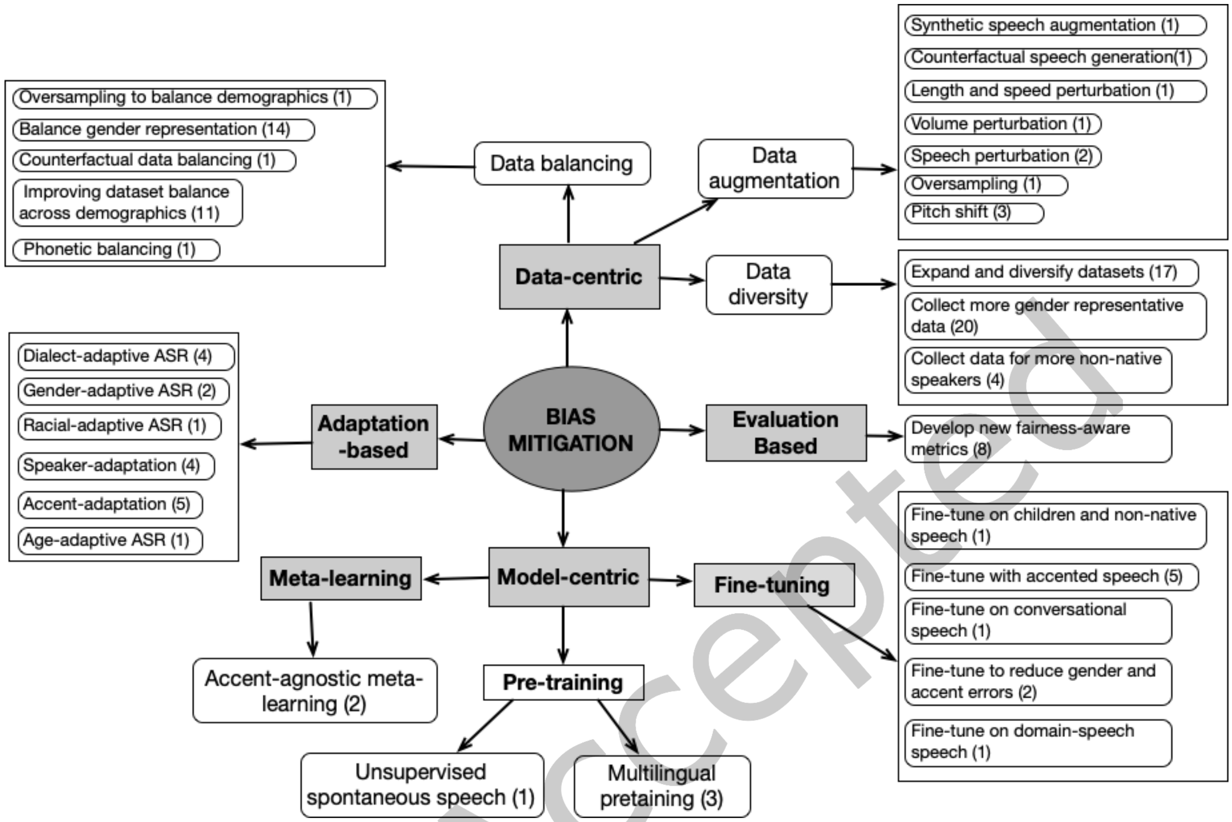

Fig. 5. Distribution of the bias mitigation strategies based on research papers on high-resourced languages. The strategies fall into four main categories: data-centric, model-centric, adaptation-based, and evaluation-based.

4.2.1 Data-centric Strategies. Across all seven studies focusing on African languages, data-centric approaches were the most commonly proposed mitigation strategies. The aim is to expand and diversify datasets to ensure equal representation of gender, age, accent, and dialect groups in training datasets [11, 31, 32, 60–62, 68]. [68] and [8] expanded and augmented datasets for Nigerian-accented English to improve accent recognition. [68] constructed a new speech corpus comprising ethnically diverse male speakers from the Yoruba, Igbo, and Hausa ethnic groups. In contrast, [8] combined Google Nigerian speech data and SautiDB to finetune NeMo's QuartzNet15x5 and XLS-R300M models. Nemo QuartzNet15x5 was pre-trained on over 3,300 hours of English speech from Mozilla Common Voice 6.1 and Multilingual LibriSpeech, incorporating speed perturbation and cutout augmentation to enhance robustness to accent variation. When creating a Kiswahili speech corpus, [80] emphasized inclusive data collection through linguistic collaboration to cover various Kiswahili dialects. They partnered with native speakers and linguists to ensure dialectal diversity, thereby reducing demographic bias at the dataset level. Similarly, [32] and [61] suggested that dataset diversity and gender balance are important in Ghanaian English and Luganda, respectively. [32] incorporated recordings from students aged 9–18 across different regions in Ghana. [61] proposed increasing female representation in training data to improve ASR performance balance, finetuning ASR models with additional Luganda speech data, and using transfer learning from related Bantu languages for model adaptation. Another approach is data augmentation, where authors propose techniques such as pitch shifting, speed perturbation, and synthetic voice generation to improve model robustness [8, 96, 100].

4.2.2 Model-centric Strategies. Four out of the seven studies on African ASR models adopted model-centric approaches, primarily through transfer learning and finetuning. These strategies involve leveraging pre-trained ASR models and adapting them to target African languages or accents using in-domain data [8, 8, 61, 98]. Finetuning NeMo's QuartzNet15x5 and Wav2Vec2.0 XLS-R300M on 4,278 Nigerian-accented English utterances from Google Nigerian speech data and SautiDB. Similarly, [32] explored finetuning Whisper V2 and Wav2Vec2.0 models using Ghanaian oral reading recordings. Despite Whisper outperforming Wav2Vec2.0 in WER, both models benefited from targeted finetuning. When evaluating gender bias in Luganda ASR, Mukiibi et al. (2022) finetuned a Kinyarwanda model on Luganda Common Voice and radio speech data. Likewise, [60] enhanced ASR accuracy for Moroccan Arabic by training Hidden Markov Models (HMMs) on dialectal speech and applying speaker adaptation techniques. Other papers have proposed dialect-specific training, where the proposal involves training separate models for different dialects or accents as an alternative approach to improve performance for specific groups [62]. These model adaptation strategies demonstrate the importance of customising pre-trained models more effectively to capture linguistic and demographic variations in African speech datasets. Another approach is accent embeddings, where accent information is incorporated into the model to help it better distinguish between different speech accents [71, 94].

4.2.3 Adaptation-based Strategies. Three studies proposed adaptation-based strategies to enhance the adaptability of ASR systems for linguistic diversity and speaker variability in African contexts. [68] introduced an accent-adaptive ASR framework targeting Nigerian-accented English. By training models on ethnically diverse speech data (Yoruba, Igbo, Hausa), the study demonstrated that adapting models to specific accent groups improved recognition accuracy and reduced misclassification across ethnic lines. [23] applied dialect-aware ASR training using a multi-dialect corpus that included Xhosa English. They proposed an Equal Accuracy Ratio (EAR) metric to reduce WER disparities across dialects, enabling fairer ASR outcomes in low-resource, multi-accent African environments. [60] focused on speaker adaptation techniques to improve ASR performance for Moroccan Arabic. By training separate models for Modern Standard Arabic (MSA) and dialectal Moroccan Arabic, the study showed that adaptation significantly boosted recognition accuracy for non-standard, dialectal speech. Model-centric techniques have also been proposed, such as speaker-adaptive techniques that dynamically adjust ASR models for different speaker characteristics [60]. Including a diverse range of speakers through speaker diversity in the training data has also been proposed as a way in which ASR models can better generalise to unseen voices. Other techniques, such as vocal tract length normalization technique, can help reduce gender bias by accounting for differences in vocal tract length between men and women.

4.2.4 Fairness-Aware Training and Evaluation. There is research that has proposed the development of new gender-aware fairness-aware ASR metrics [19], fairness-aware model evaluation, such as the development of new metrics to measure disparities across demographic groups [1, 24, 31], and fairness-aware training techniques that aim to minimize bias during the training process, ensuring that the model performs equally well for all groups [6, 24, 81, 86]. In high-resource languages, ASR bias has been studied extensively. Research indicates that commercial ASR systems from companies such as Google, Amazon, and Microsoft frequently exhibit disparities in error rates across different genders, accents, and age groups. These studies highlight systemic biases, such as higher error rates for one gender and non-native accents. Despite significant efforts to develop fairness-aware ASR training in these languages, similar initiatives are mainly missing in Africa. Most African ASR research has focused on expanding datasets or applying transfer learning, rather than integrating advanced bias-aware learning techniques, such as adversarial debiasing or fairness-driven loss functions.

When ASR systems consistently perform poorly for certain speaker groups, they risk reinforcing existing inequalities and making technology less accessible for marginalized communities. For example, one study reported that a Luganda ASR model trained on gender-imbalanced data resulted in a 70.6% word error rate for female speakers compared to 53.5% for male speakers, which demonstrates how dataset imbalances can distort performance [61]. In another case, mainstream ASR services like Google Speech Recognition struggled with Nigerian-accented English, exhibiting a high word error rate of 44.2%, which serves as a strong reminder that current systems are not well-adapted to Africa's rich linguistic diversity [8]. If these biases are not addressed, ASR technology risks continuing to favour dominant speaker groups while abandoning underrepresented populations, which eventually limits the effectiveness of these tools in real-world applications.

Despite the insights obtained from this study, several challenges remain. The limited availability of research on ASR bias in African languages makes it hard to establish comprehensive benchmarks for fairness evaluation. While gender, accent, and dialect biases have been explored, there is a clear gap when it comes to understanding how ASR models perform across different age groups and races. Future work should focus on expanding publicly available African speech datasets that are gender-balanced, demographically diverse, and dialect-inclusive. As demonstrated in the Kiswahili corpus development study [80], collaborating with linguists and local communities can significantly improve dataset inclusivity. However, simply improving the data is not enough. Future studies should also explore algorithmic fairness techniques, such as fairness-aware loss functions [24, 44] and adversarial debiasing strategies [81], which have already shown promise in high-resource ASR research.

RQ2: *What strategies have been proposed to mitigate these biases and promote inclusion in ASR systems?*
Summary for RQ2: *Data-centric strategies were the dominant approach, appearing in all 7 African ASR studies, and were primarily applied to mitigate accent and dialect bias. Model-centric techniques such as finetuning and transfer learning were used in 6 studies (86%), often targeting gender and accent disparities. Adaptation-based methods were less common (43%), with a primary focus on dialect variation. In particular, no mitigation strategies were found that addressed age or racial biases. These findings highlight that while current efforts in African ASR research strongly emphasize accent, dialect, and gender fairness, other critical forms of bias remain largely unaddressed, indicating significant gaps and opportunities for future work.*

5 Discussion and Research Directions

Our review of ASR bias in low-resource African languages reveals that imbalanced training data is a key factor driving disparities in model performance. This imbalance leads to the most commonly studied biases: accent, dialect, and gender, which result in higher word error rates for underrepresented speaker groups. Differences in word error rates hinder and limit the effectiveness of using ASR models in real-world applications.

Our review found that imbalanced training data mainly drives bias in ASR models for low-resource African languages. The most commonly studied biases (accent, dialect, and gender) led to higher word error rates for underrepresented groups, ultimately limiting the inclusivity of these technologies. Although several bias mitigation strategies have been proposed, most focus on data enhancement and demographic balancing. Fairness-aware modelling techniques specifically designed for African languages have seen limited exploration.

Although numerous bias mitigation strategies have been proposed, most focus on data enhancement and demographic balancing. These are important first steps, and our review highlights a range of techniques to achieve them, including adjusting model hyperparameters for unknown dialects, adaptive feature-wise transformation, joint conditioning on dialect and internal representation, introducing unseen dialect classes for generalization, zero-shot adaptation, and self-supervised learning. Similarly, several approaches aim to improve dialect and accent handling through accent embeddings, incorporating accent feature extraction modules, minimising mutual information between native and non-native speech embeddings, utilising accent-agnostic ASR architectures, employing dialect-aware language models, and integrating diverse linguistic features. Model structure and processing techniques, such as personalised speech models, autoencoder-based approaches, spectrogram transformation, and pre-training with spontaneous speech, have also been explored.

However, there is a significant gap in exploring fairness-based modeling techniques specifically designed for African languages. This contrasts with extensive research on ASR bias in high-resource languages. Studies in high-resource settings have consistently shown that commercial ASR systems from major technology companies (e.g., Google, Amazon, and Microsoft) exhibit substantial disparities in error rates across various demographic groups, including gender, accent, and age. These systems often demonstrate systemic biases, such as higher error rates for female speakers and non-native accents. While significant efforts have been made to develop fairness-aware ASR training in these high-resource languages, similar initiatives are limited in the context of African languages. Instead, African ASR research has predominantly focused on expanding datasets or applying transfer learning rather than integrating advanced bias-aware learning techniques. These advanced techniques, such as fairness-aware loss functions [24, 44] and adversarial debiasing strategies [81], have shown promise in mitigating bias in high-resource ASR systems, but their application to African languages remains limited.

The consequences of ASR systems that consistently underperform for certain speaker groups are significant. These systems risk reinforcing existing inequalities, perpetuating marginalization, and hindering technology adoption among minority communities. For instance, the Luganda ASR model study [61] demonstrated how dataset imbalances can distort performance, with a reported 70.6% word error rate for female speakers compared to 53.5% for male speakers. Similarly, the difficulties encountered by mainstream ASR services, such as Google Speech Recognition, with Nigerian-accented English resulted in a high 44.2% word error rate [8], highlighting the challenge of adapting current systems to Africa's rich linguistic diversity. Without addressing these systemic biases, ASR technology may inadvertently favour dominant speaker groups, further marginalizing underrepresented populations and limiting the technology's overall effectiveness and utility.

5.1 Research Gaps

Despite the insights gained from existing research on biases and their mitigation for ASR models, several challenges and research gaps remain. The limited availability of research on ASR bias in African languages makes it difficult to establish comprehensive benchmarks for fairness evaluation. Although gender, accent and dialect biases have been explored, there is a notable lack of understanding regarding how ASR models perform in different age groups and races within the African context. Future research directions are needed to address these gaps.

Insufficient coverage of bias studies for African languages. Our analysis revealed that only 7 of the 55 selected studies explicitly investigated biases in ASR for low-resource African languages, compared to 51 that addressed high-resource languages such as English, French, Spanish and Dutch (Figure 3). However, the limitation is not only in quantity but also in scope. African-focused studies focused mainly on accent and dialect biases, such as in Nigerian-accented English [8, 68], Kiswahili dialects [80], Moroccan Arabic dialects [60], and South African English varieties including Xhosa [23]. Gender bias was explicitly studied only in Luganda [61], while age and racial biases were completely absent. These findings show that existing studies do not provide comprehensive information on the full spectrum of bias challenges in African ASR. Addressing this gap requires not only more research work but also broader coverage that systematically evaluates multiple demographic dimensions, including age, race, and their intersection with gender, accent, and dialect. In addition, there is a critical need for publicly available African speech datasets that are gender-balanced, demographically diverse, and inclusive of various dialects. The Kiswahili corpus development study [80] provides a valuable model that demonstrates how collaboration with linguists and local communities can enhance the inclusion and representation of the datasets.

Dominance of data-centric strategies with limited exploration of advanced techniques. Our review found that data-centric approaches, such as dataset expansion, diversity enhancement and data enhancement, were the most prevalent bias mitigation strategies in the seven African language studies (100%). For example, [68] and [8] primarily focused on expanding and diversifying the English datasets acquired by Nigeria. Similarly, [80] highlighted the importance of collaborative data collection to create datasets that include Kiswahili dialects. However, more advanced model-centric approaches, such as fairness-aware model architectures and adversarial debiasing [67], remain unexplored in African contexts, despite their popularity in high-resource languages [81]. This gap highlights the need for research that extends beyond basic dataset improvements to sophisticated bias mitigation methods tailored for low-resource settings.

Lack of specialized fairness evaluation metrics in African ASR research. African ASR studies primarily use standard metrics, such as Word Error Rate (WER) or Character Error Rate (CER), for bias evaluation. In contrast, high-resource language studies frequently propose specialized fairness metrics to evaluate and quantify bias [34]. Among the African studies reviewed, only [23] proposed a fairness-based evaluation metric. Other studies did not explicitly develop or apply fairness-specific metrics. The absence of specialised fairness evaluation frameworks limits a comprehensive understanding, identification, and systematic reduction of biases in ASR for African languages. Beyond improving data, future studies should focus on developing and applying algorithmic fairness techniques tailored to the specific challenges of African languages. This includes exploring the effectiveness of fairness-aware loss functions [24, 44] and adversarial debiasing strategies [81], which have shown promise in mitigating bias in high-resource ASR research but require further investigation in low-resource African language settings.

Insufficient multi-dimensional and intersectional bias analysis. Our findings suggest that African ASR research frequently examines biases in isolation, primarily focusing on gender, accent, or dialect biases separately. The research in [68] acknowledged the underrepresentation of female voices but did not quantitatively explore the intersection between gender and accent bias. Intersectional analyzes have become increasingly significant in high-resource language research [22, 74], yet remain largely unaddressed in African ASR contexts. This involves examining how various factors, including gender, accent, dialect, age, and socioeconomic status, impact ASR performance.

Future research should focus on creating datasets with balanced and explicitly annotated metadata across multiple demographic axes (e.g., gender, age, dialect, and region). This would allow the evaluation of ASR models at specific intersections, such as "young, female, Xhosa-accented English speakers", to uncover nuanced and hidden biases that current single-axis evaluations overlook.

Lack of longitudinal bias evaluation studies. As a further area of research, longitudinal studies are needed to understand how ASR bias evolves, mainly as models are updated and new data from African contexts are introduced. This research can also focus on understanding linguistically similar African languages and how this introduces biases in ASR models [80]. This will help ensure that fairness gains are maintained and that new biases are not introduced inadvertently.

6 Conclusion

In this study, we examined biases in ASR models for low-resource African languages, with a specific focus on accent, dialect, and gender. Our analysis reveals that these biases are primarily due to data imbalances and limited linguistic diversity within training datasets, resulting in high word error rates for underrepresented speaker groups. For example, ASR systems face challenges with various English accents, including Nigerian, Xhosa, and Ghanaian, as well as with different Moroccan Arabic dialects. We also observed significant gender bias, as exemplified in languages such as Luganda, where the predominance of male voices in training data leads to performance disparities.

The predominant approach to mitigating these biases in African languages has been data-driven, emphasizing strategies such as expanding the diversity of training datasets, employing data augmentation techniques, and applying transfer learning from high-resource languages. However, our review indicates a notable gap in exploring fairness-aware modelling strategies. Techniques such as bias-aware loss functions, adversarial debiasing, and speaker-adaptive models, while promising, remain underexplored in this context. This highlights a critical need for future research to pursue a two-pronged approach: enhancing dataset quality and integrating fairness-focused algorithms to achieve more comprehensive and practical solutions.

Although existing studies on African languages have predominantly focused on accent and dialect biases, there is limited research on gender bias, age bias, and other potential sources of disparity. Furthermore, the limited number of African languages examined to date highlights the need for further research to understand biases and develop mitigation strategies for African languages. Expanding the scope of research to encompass a broader range of African languages and a more diverse spectrum of biases will provide a more comprehensive understanding of these issues and facilitate the development of fairer and more robust ASR technologies across the continent.

Acknowledgments

This work was funded by the Gender and Responsible Artificial Intelligence Network (GRAIN), a consortium comprising Initiative Prospective Agricole et Rurale (IPAR), Centre for the Study of the Economy in Africa (CSEA), and Sunbird AI.

References

[1] Martine Adda-Decker and Lori Lamel. 2005. Do speech recognizers prefer female speakers?. In *Interspeech*. 2205–2208.

[2] Ife Adebara and Muhammad Abdul-Mageed. 2022. Towards Afrocentric NLP for African Languages: Where We Are and Where We Can Go. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 3814–3841. doi:10.18653/v1/2022.acl-long.265

[3] Benjamin Akera, Jonathan Mukiibi, Lydia Sanyu Naggayi, Claire Babirye, Isaac Owomugisha, Solomon Nsumba, Joyce Nakatumba-Nabende, Engineer Bainomugisha, Ernest Mwebaze, and John Quinn. 2022. Machine translation for african languages: Community creation of datasets and models in uganda. In *3rd Workshop on African Natural Language Processing*.

[4] Guilherme Andrade, Luiz Nery, Fabricio Benevenuto, Flavio Figueiredo, and Savvas Zannettou. 2024. Comprehensive View of the Biases of Toxicity and Sentiment Analysis Methods Towards Utterances with African American English Expressions. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil, 1–10. doi:10.5753/webmedia.2024.243106

[5] Surakshya Aryal, Mikel K. Ngueajio, Saurav Keshari Aryal, and Gloria Washington. 2023. Hey, siri! why are you biased against women? (student abstract). In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'23/IAAI'23/EAAI'23)*. AAAI Press, Article 1868, 2 pages. doi:10.1609/aaai.v37i13.26937

[6] Giuseppe Attanasio, Beatrice Savoldi, Dennis Fucci, and Dirk Hovy. 2024. Twists, Humps, and Pebbles: Multilingual Speech Recognition Models Exhibit Gender Performance Gaps. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 21318–21340. doi:10.18653/v1/2024.emnlp-main.1188

[7] Ebbie Awino, Lilian Wanzare, Lawrence Muchemi, Barack Wamkaya Wanjawa, Edward Ombui, Florence Indede, Owen McOnyango, and Benard Okal. 2022. Phonemic Representation and Transcription for Speech to Text Applications for Under-resourced Indigenous African Languages: The Case of Kiswahili. *CoRR* abs/2210.16537 (2022). doi:10.48550/ARXIV.2210.16537

[8] Oreoluwa Boluwatife Babatunde, Emmanuel Akeweje, Sharon Ibejih, Victor Tolulope Olufemi, and Sakinat Oluwabukonla Folorunso. 2023. Automatic speech recognition for Nigerian-accented English. In *Deep Learning Indaba 2023*.

[9] Vivek Bhardwaj, Mohamed Tahar Ben Othman, Vinay Kukreja, Youcef Belkhier, Mohit Bajaj, B. Srikanth Goud, Ateeq Ur Rehman, Muhammad Shafiq, and Habib Hamam. 2022. Automatic Speech Recognition (ASR) Systems for Children: A Systematic Literature Review. *Applied Sciences* 12, 9 (2022). doi:10.3390/app12094419

[10] Amanda Cercas Curry and Verena Rieser. 2018. #MeToo Alexa: How Conversational Systems Respond to Sexual Harassment. In *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing*, Mark Alfano, Dirk Hovy, Margaret Mitchell, and Michael Strube (Eds.). Association for Computational Linguistics, New Orleans, Louisiana, USA, 7–14. doi:10.18653/v1/W18-0802

[11] Anastasiia Chizhikova, Hannah Billinghurst, Michelle Elizabeth, Shehenaz Hossain, Ajinkya Kulkarni, Gaël Guibon, and Miguel Couceiro. 2024. Factorizing Gender Bias in Automatic Speech Recognition for Mexican Spanish. (2024).

[12] June Choe, Yiran Chen, May Pik Yu Chan, Aini Li, Xin Gao, and Nicole Holliday. 2022. Language-specific Effects on Automatic Speech Recognition Errors for World Englishes. In *Proceedings of the 29th International Conference on Computational Linguistics*, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na (Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 7177–7186. https://aclanthology.org/2022.coling-1.628/

[13] Amit Das, Preethi Jyothi, and Mark Hasegawa-Johnson. 2016. Automatic Speech Recognition Using Probabilistic Transcriptions in Swahili, Amharic, and Dinka.. In *INTERSPEECH*. 3524–3528.

[14] Miguel Del Río, Corey Miller, Ján Profant, Jennifer Drexler-Fox, Quinn Mcnamara, Nishchal Bhandari, Natalie Delworth, Ilya Pirkin, Migüel Jetté, Shipra Chandra, et al. 2023. Accents in speech recognition through the lens of a world englishes evaluation set. *Research in Language* 21, 3 (2023), 225–244.

[15] Rachel Dorn. 2019. Dialect-specific models for automatic speech recognition of African American Vernacular English. In *Proceedings of the Student Research Workshop Associated with RANLP 2019*. 16–20.

[16] Bonaventure F. P. Dossou and Chris Chinenye Emezue. 2021. OkwuGbé: End-to-End Speech Recognition for Fon and Igbo. In *Proceedings of the Fifth Workshop on Widening Natural Language Processing*, Erika Varis, Ryan Georgi, Alicia Tsai, Antonios Anastasopoulos, Kyathi Chandu, Xanda Schofield, Surangika Ranathunga, Haley Lepp, and Tirthankar Ghosal (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 1–4. https://aclanthology.org/2021.winlp-1.1/

[17] Moayad Elamin, Yonas Chanie, Paul Ewuzie, and Samuel Rutunda. 2023. Multilingual Automatic Speech Recognition for Kinyarwanda, Swahili, and Luganda: Advancing ASR in Select East African Languages. In *4th Workshop on African Natural Language Processing*.

[18] Sidney Evaldo Leal, Arnaldo Candido Junior, Ricardo Marcacini, Edresson Casanova, Odilon Gonçalves, Anderson Silva Soares, Rodrigo Freitas Lima, Lucas Rafael Stefanel Gris, and Sandra Aluísio. 2025. MuPe Life Stories Dataset: Spontaneous Speech in Brazilian Portuguese with a Case Study Evaluation on ASR Bias against Speakers Groups and Topic Modeling. In *Proceedings of the 31st International Conference on Computational Linguistics*, Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert (Eds.). Association for Computational Linguistics, Abu Dhabi, UAE, 6076–6087. https://aclanthology.org/2025.coling-main.407/

[19] Siyuan Feng, Bence Mark Halpern, Olya Kudina, and Odette Scharenborg. 2024. Towards inclusive automatic speech recognition. *Computer Speech & Language* 84 (2024), 101567.

[20] Siyuan Feng, Olya Kudina, Bence Mark Halpern, and Odette Scharenborg. 2021. Quantifying bias in automatic speech recognition. *arXiv preprint arXiv:2103.15122* (2021).

[21] Dennis Fucci, Marco Gaido, Matteo Negri, Mauro Cettolo, and Luisa Bentivogli. 2023. No Pitch Left Behind: Addressing Gender Unbalance in Automatic Speech Recognition through Pitch Manipulation. In *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 1–8.

[22] Marcio Fuckner, Sophie Horsman, Pascal Wiggers, and Iskaj Janssen. 2023. Uncovering Bias in ASR Systems: Evaluating Wav2vec2 and Whisper for Dutch speakers. In *2023 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. IEEE, 146–151.

[23] Heting Gao, Xiaoxuan Wang, Sunghun Kang, Rusty Mina, Dias Issa, John Harvill, Leda Sari, Mark Hasegawa-Johnson, and Chang D Yoo. 2022. Seamless equal accuracy ratio for inclusive CTC speech recognition. *Speech Communication* 136 (2022), 76–83.

[24] Mahault Garnerin, Solange Rossato, and Laurent Besacier. 2019. Gender representation in French broadcast corpora and its impact on ASR performance. In *Proceedings of the 1st international workshop on AI for smart TV content production, access and delivery*. 3–9.

[25] Mahault Garnerin, Solange Rossato, and Laurent Besacier. 2021. Investigating the Impact of Gender Representation in ASR Training Data: a Case Study on Librispeech. In *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*, Marta Costa-jussa, Hila Gonen, Christian Hardmeier, and Kellie Webster (Eds.). Association for Computational Linguistics, Online, 86–92. doi:10.18653/v1/2021.gebnlp-1.10

[26] Elodie Gauthier, Laurent Besacier, Sylvie Voisin, Michael Melese, and Uriel Pascal Elingui. 2016. Collecting resources in sub-saharan african languages for automatic speech recognition: a case study of wolof. In *10th Language Resources and Evaluation Conference (LREC 2016)*.

[27] Reem Gody and David Harwath. 2023. Unsupervised fine-tuning data selection for asr using self-supervised speech models. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 1–5.

[28] Calbert Graham and Nathan Roll. 2024. Evaluating OpenAI's Whisper ASR: Performance analysis across diverse accents and speaker traits. *JASA Express Letters* 4, 2 (2024).

[29] Alexander Gutkin, Isin Demirsahin, Oddur Kjartansson, Clara E. Rivera, and Kólá Túbòsún. 2020. Developing an Open-Source Corpus of Yoruba Speech. In *Proc. of Interspeech 2020*. October 25–29, Shanghai, China, 2020., 404–408. http://dx.doi.org/10.21437/Interspeech.2020-1096

[30] Camille Harris, Chijioke Mgbahurike, Neha Kumar, and Diyi Yang. 2024. Modeling Gender and Dialect Bias in Automatic Speech Recognition. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (Eds.). Association for Computational Linguistics, Miami, Florida, USA, 15166–15184. doi:10.18653/v1/2024.findings-emnlp.890

[31] Foyzul Hassan, Mohammad Saiful Alam Khan, Mohammed Rokibul Alam Kotwal, and Mohammad Nurul Huda. 2012. Gender independent Bangla automatic speech recognition. In *2012 international conference on informatics, electronics & vision (ICIEV)*. IEEE, 144–148.

[32] Owen Henkel, Hannah Horne-Robinson, Libby Hills, Bill Roberts, and Josh McGrane. 2025. Supporting Literacy Assessment in West Africa: Using State-of-the-Art Speech Models to Assess Oral Reading Fluency. *International Journal of Artificial Intelligence in Education* (2025), 1–22.

[33] Aaricia Herygers, Vass Verkhodanova, Matt Coler, Odette Scharenborg, and Munir Georges. 2023. Bias in Flemish Automatic Speech Recognition. In *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2023*, Christoph Draxler (Ed.). TUDpress, Dresden, 158–165. https://www.essv.de/pdf/2023_158_165.pdf

[34] Samuel Hollands, Daniel Blackburn, and Heidi Christensen. 2022. Evaluating the performance of state-of-the-art ASR systems on non-native English using corpora with extensive language background variation. In *Interspeech 2022: Proceedings of the Annual Conference of the International Speech Communication Association*. International Speech Communication Association (ISCA), 3958–3962.

[35] Catherine Ikae and Mascha Kurpicz-Briki. 2024. Current State-of-the-Art of Bias Detection and Mitigation in Machine Translation for African and European Languages: a Review. *arXiv preprint arXiv:2410.21126* (2024).

[36] Saki Imai, Tahiya Chowdhury, and Amanda J. Stent. 2025. Evaluating Open-Source ASR Systems: Performance Across Diverse Audio Conditions and Error Correction Methods. In *Proceedings of the 31st International Conference on Computational Linguistics*, Owen Rambow, Leo Wanner, Marianna Apidianaki, Hend Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert (Eds.). Association for Computational Linguistics, Abu Dhabi, UAE, 5027–5039. https://aclanthology.org/2025.coling-main.336/

[37] Jazette Johnson, Lucretia Williams, Jaye Nias, Saurav K. Aryal, and Gloria Washington. 2025. Centering Black Voices: Lessons Learned and Reflections from a Large-Scale AAVE Data Collection at a Historically Black University. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 658, 7 pages. doi:10.1145/3706599.3706669

[38] Kartik Khandelwal, Preethi Jyothi, Abhijeet Awasthi, and Sunita Sarawagi. 2020. Black-box adaptation of ASR for accented speech. *arXiv preprint arXiv:2006.13519* (2020).

[39] Ussen Kimanuka, Ciira wa Maina, and Osman Büyük. 2024. Speech recognition datasets for low-resource Congolese languages. *Data in Brief* 52 (2024), 109796.

[40] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, 2004 (2004), 1–26.

[41] Alexander R Kivaisi, Qingjie Zhao, and Jimmy T Mbelwa. 2023. Swahili speech dataset development and improved pre-training method for spoken digit recognition. *ACM Transactions on Asian and Low-Resource Language Information Processing* 22, 7 (2023), 1–24.

[42] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R Rickford, Dan Jurafsky, and Sharad Goel. 2020. Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* 117, 14 (2020), 7684–7689.

[43] Ajinkya Kulkarni, Anna Tokareva, Rameez Qureshi, and Miguel Couceiro. 2024. The Balancing Act: Unmasking and Alleviating ASR Biases in Portuguese. In *Proceedings of the Fourth Workshop on Language Technology for Equality, Diversity, Inclusion*, Bharathi Raja Chakravarthi, Bharathi B, Paul Buitelaar, Thenmozhi Durairaj, György Kovács, and Miguel Ángel García Cumbreras (Eds.). Association for Computational Linguistics, St. Julian's, Malta, 31–40. https://aclanthology.org/2024.ltedi-1.4/

[44] Chunxi Liu, Michael Picheny, Leda Sarı, Pooja Chitkara, Alex Xiao, Xiaohui Zhang, Mark Chou, Andres Alvarado, Caner Hazirbas, and Yatharth Saraf. 2022. Towards measuring fairness in speech recognition: Casual conversations dataset transcriptions. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6162–6166.

[45] Nóra Ni Loideain and Rachel Adams. 2020. From Alexa to Siri and the GDPR: The gendering of Virtual Personal Assistants and the role of Data Protection Impact Assessments. *Computer Law & Security Review* 36 (2020), 105366. doi:10.1016/j.clsr.2019.105366

[46] Liam Lonergan, Mengjie Qian, Neasa Ní Chiaráin, Christer Gobl, and Ailbhe Ní Chasaide. 2024. Low-resource speech recognition and dialect identification of Irish in a multi-task framework. In *The Speaker and Language Recognition Workshop (Odyssey 2024)*. 67–73. doi:10.21437/odyssey.2024-10

[47] Liam Lonergan, Mengjie Qian, Neasa Ní Chiaráin, Christer Gobl, and Ailbhe Ní Chasaide. 2023. Towards Dialect-inclusive Recognition in a Low-resource Language: Are Balanced Corpora the Answer?. In *Interspeech 2023*. 5082–5086. doi:10.21437/Interspeech.2023-2528

[48] Liam Lonergan, Ibon Saratxaga, John Sloan, Oscar Maharg Bravo, Mengjie Qian, Neasa Ní Chiaráin, Christer Gobl, and Ailbhe Ní Chasaide. 2025. Fotheidil: an Automatic Transcription System for the Irish Language. In *Proceedings of the 5th Celtic Language Technology Workshop*, Brian Davis, Theodorus Fransen, Elaine Uí Dhonnchadha, and Abigail Walsh (Eds.). International Committee on

Computational Linguistics, Abu Dhabi [Virtual Workshop], 35–45. https://aclanthology.org/2025.cltw-1.4/

[49] Koena Ronny Mabokela, Turgay Celik, and Mpho Raborife. 2023. Multilingual Sentiment Analysis for Under-Resourced Languages: A Systematic Review of the Landscape. *IEEE Access* 11 (2023), 15996–16020. doi:10.1109/ACCESS.2022.3224136

[50] Nina Markl. 2022. Language variation and algorithmic bias: understanding algorithmic bias in British English automatic speech recognition. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 521–534.

[51] Joshua L Martin and Kevin Tang. 2020. Understanding Racial Disparities in Automatic Speech Recognition: The Case of Habitual" be".. In *Interspeech*. 626–630.

[52] Joshua L Martin and Kelly Elizabeth Wright. 2023. Bias in automatic speech recognition: The case of African American language. *Applied Linguistics* 44, 4 (2023), 613–630.

[53] Yen Meng, Yi-Hui Chou, Andy T Liu, and Hung-yi Lee. 2022. Don't speak too fast: The impact of data bias on self-supervised speech models. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 3258–3262.

[54] Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "I don't Think These Devices are Very Culturally Sensitive."—Impact of Automated Speech Recognition Errors on African Americans. *Frontiers in Artificial Intelligence* 4 (2021), 725911.

[55] Raghav Menon, Armin Saeb, Hugh Cameron, William Kibira, John Quinn, and Thomas Niesler. 2017. Radio-browsing for developmental monitoring in Uganda. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5795–5799.

[56] Josh Meyer, Lindy Rauchenstein, Joshua D. Eisenberg, and Nicholas Howell. 2020. Artie Bias Corpus: An Open Dataset for Detecting Demographic Bias in Speech Applications. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 6462–6468. https://aclanthology.org/2020.lrec-1.796/

[57] Naira Abdou Mohamed, Imade Benelallam, Anass Allak, and Kamel Gaanoun. 2023. Multilingual Speech Recognition Initiative for African Languages. (2023).

[58] Jama Hussein Mohamud, Lloyd Acquaye Thompson, Aissatou Ndoye, and Laurent Besacier. 2021. Fast development of asr in african languages using self supervised speech representation learning. *arXiv preprint arXiv:2103.08993* (2021).

[59] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, and t PRISMA Group*. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151, 4 (2009), 264–269.

[60] Bezoui Mouaz, Beni Hssane Abderrahim, and Elmoutaouakkil Abdelmajid. 2019. Speech recognition of moroccan dialect using hidden Markov models. *Procedia Computer Science* 151 (2019), 985–991.

[61] Jonathan Mukiibi, Andrew Katumba, Joyce Nakatumba-Nabende, Ali Hussein, and Joshua Meyer. 2022. The Makerere Radio Speech Corpus: A Luganda Radio Corpus for Automatic Speech Recognition. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association, Marseille, France, 1945–1954. https://aclanthology.org/2022.lrec-1.208/

[62] Jonghwan Na, Yeseul Park, and Bowon Lee. 2024. A Comparative Study on the Biases of Age, Gender, Dialects, and L2 speakers of Automatic Speech Recognition for Korean Language. In *2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 1–6. doi:10.1109/APSIPAASC63619.2025.10848815

[63] Eduardo Nacimiento-García, Holi Sunya Díaz-Kaas-Nielsen, and Carina S González-González. 2024. Gender and Accent Biases in AI-Based Tools for Spanish: A Comparative Study between Alexa and Whisper. *Applied Sciences* 14, 11 (2024), 4734.

[64] Joyce Nakatumba-Nabende, Claire Babirye, Peter Nabende, Jeremy Francis Tusubira, Jonathan Mukiibi, Eric Peter Wairagala, Chodrine Mutebi, Tobius Saul Bateesa, Alvin Nahabwe, Hewitt Tusiime, et al. 2024. Building Text and Speech Benchmark Datasets and Models for Low-Resourced East African Languages: Experiences and Lessons. *Applied AI Letters* 5, 2 (2024), e92.

[65] Joyce Nakatumba-Nabende, Jonathan Mukiibi, Tobius Saul Bateesa, Sudi Murindanyi, Andrew Katumba, and Chodrine Mutebi. 2024. Machine Learning Analysis of Radio Data to Uncover Community Perceptions on the Ebola Outbreak in Uganda. *ACM Journal on Computing and Sustainable Societies* (2024).

[66] Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. 2020. Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Online, 2144–2160. doi:10.18653/v1/2020.findings-emnlp.195

[67] Mikel K. Ngueajio and Gloria Washington. 2022. Hey ASR System! Why Aren't You More Inclusive? Automatic Speech Recognition Systems' Bias and Proposed Bias Mitigation Techniques. A Literature Review. In *HCI International 2022 – Late Breaking Papers: Interacting with EXtended Reality and Artificial Intelligence: 24th International Conference on Human-Computer Interaction, HCII 2022, Virtual Event, June 26 – July 1, 2022, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 421–440. doi:10.1007/978-3-031-21707-4_30

[68] Francisca O Oladipo, Rahmon A Habeeb, Abraham E Musa, Chinecherem Umezuruike, and Ohieku Andrew Adeiza. 2021. Automatic Speech Recognition and Accent Identification of Ethnically Diverse Nigerian English Speakers. *International Journal of Applied Information Systems* 12, 36 (2021), 41–48.

[69] Tobi Olatunji, Tejumade Afonja, Bonaventure F. P. Dossou, Atnafu Lambebo Tonja, Chris Chinenye Emezue, Amina Mardiyyah Rufai, and Sahib Singh. 2023. AfriNames: Most ASR Models "Butcher" African Names. In *Interspeech 2023*. 5077–5081. doi:10.21437/Interspeech.2023-2122

[70] Abigail Oppong. 2023. Building a participatory data design approach to examine gender bias in english-twi machine translation. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6.

[71] Abraham Owodunni, Aditya Yadavalli, Chris Emezue, Tobi Olatunji, and Clinton Mbataku. 2024. AccentFold: A Journey through African Accents for Zero-Shot ASR Adaptation to Target Accents. In *Findings of the Association for Computational Linguistics: EACL 2024*, Yvette Graham and Matthew Purver (Eds.). Association for Computational Linguistics, St. Julian's, Malta, 2146–2161. https://aclanthology.org/2024.findings-eacl.142/

[72] Emma Pair, Nikitha Vicas, Ann M Weber, Valerie Meausoone, James Zou, Amos Njuguna, and Gary L Darmstadt. 2021. Quantification of gender bias and sentiment toward political leaders over 20 years of Kenyan news using natural language processing. *Frontiers in Psychology* 12 (2021), 712646.

[73] Kacper Radzikowski, Le Wang, Osamu Yoshie, and Robert Nowak. 2021. Accent modification for speech recognition of non-native speakers using neural style transfer. *EURASIP Journal on Audio, Speech, and Music Processing* 2021, 1 (2021), 11.

[74] Leda Sarı, Mark Hasegawa-Johnson, and Chang D Yoo. 2021. Counterfactually fair automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 29 (2021), 3515–3525.

[75] Majdi Sawalha and Mohammad Abu Shariah. 2013. The effects of speakers' gender, age, and region on overall performance of Arabic automatic speech recognition systems using the phonetically rich and balanced Modern Standard Arabic speech corpus. In *Proceedings of the 2nd Workshop of Arabic Corpus Linguistics WACL-2*. Leeds.

[76] Dana Serditova, Kevin Tang, and Jochen Steffens. 2025. Automatic Speech Recognition Biases in Newcastle English: an Error Analysis. arXiv:2506.16558 [cs.CL] https://arxiv.org/abs/2506.16558

[77] Walelign Tewabe Sewunetie, Atnafu Lambebo Tonja, Tadesse Destaw Belay, Hellina Hailu Nigatu, Gashaw Kidanu, Zewdie Mossie, Hussien Seid, Eshete Derb, and Seid Muhie Yimam. 2024. Evaluating gender bias in machine translation for low-resource languages. In *5th Workshop on African Natural Language Processing*.

[78] Sunny Shrestha and Sanchari Das. 2022. Exploring gender biases in ML and AI academic research through systematic literature review. *Frontiers in artificial intelligence* 5 (2022), 976838.

[79] Claytone Sikasote, Kalinda Siaminwe, Stanly Mwape, Bangiwe Zulu, Mofya Phiri, Martin Phiri, David Zulu, Mayumbo Nyirenda, and Antonios Anastasopoulos. 2023. Zambezi Voice: A Multilingual Speech Corpus for Zambian Languages. In *Interspeech 2023*. 3984–3988. doi:10.21437/Interspeech.2023-1979

[80] Kathleen Siminyu, Kibibi Mohamed Amran, Abdulrahman Ndegwa Karatu, Mnata Resani, Mwimbi Makobo Junior, Rebecca Ryakitimbo, and Britone Mwasaru. 2022. Corpus Development of Kiswahili Speech Recognition Test and Evaluation sets, Preemptively Mitigating Demographic Bias Through Collaboration with Linguists. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*. 13–19.

[81] Anna Smerdiagina. 2024. Lost in transcription: Experimental findings on ethnic and age biases in AI systems. *Junior Management Science (JUMS)* 9, 3 (2024), 1591–1608.

[82] Karolina Stańczak and Isabelle Augenstein. 2021. A Survey on Gender Bias in Natural Language Processing. *ArXiv* abs/2112.14168 (2021). https://api.semanticscholar.org/CorpusID:245537923

[83] Tony Sun, Andrew Gaut, Shirlyn Tang, Yuxin Huang, Mai ElSherief, Jieyu Zhao, Diba Mirza, Elizabeth Belding, Kai-Wei Chang, and William Yang Wang. 2019. Mitigating Gender Bias in Natural Language Processing: Literature Review. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Anna Korhonen, David Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, Florence, Italy, 1630–1640. doi:10.18653/v1/P19-1159

[84] Martha Yifiru Tachbelie, Solomon Teferra Abate, and Tanja Schultz. 2020. Development of Multilingual ASR Using GlobalPhone for Less-Resourced Languages: The Case of Ethiopian Languages.. In *INTERSPEECH*. 1032–1036.

[85] Rachael Tatman. 2017. Gender and Dialect Bias in YouTube's Automatic Captions. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, Dirk Hovy, Shannon Spruit, Margaret Mitchell, Emily M. Bender, Michael Strube, and Hanna Wallach (Eds.). Association for Computational Linguistics, Valencia, Spain, 53–59. doi:10.18653/v1/W17-1606

[86] Rachael Tatman and Conner Kasten. 2017. Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions.. In *Interspeech*. 934–938.

[87] Melissa Torgbi, Andrew Clayman, Jordan J. Speight, and Harish Tayyar Madabushi. 2025. Adapting Whisper for Regional Dialects: Enhancing Public Services for Vulnerable Populations in the United Kingdom. In *Proceedings of the 12th Workshop on NLP for Similar Languages, Varieties and Dialects*, Yves Scherrer, Tommi Jauhiainen, Nikola Ljubešić, Preslav Nakov, Jorg Tiedemann, and Marcos Zampieri (Eds.). Association for Computational Linguistics, Abu Dhabi, UAE, 29–38. https://aclanthology.org/2025.vardial-1.4/

[88] Anja Virkkunen, Aku Rouhe, Nhan Phan, and Mikko Kurimo. 2023. Finnish parliament ASR corpus: Analysis, benchmarks and statistics. *Language Resources and Evaluation* 57, 4 (2023), 1645–1670.

[89] Eric Peter Wairagala, Jonathan Mukiibi, Jeremy Francis Tusubira, Claire Babirye, Joyce Nakatumba-Nabende, Andrew Katumba, and Ivan Ssenkungu. 2022. Gender bias evaluation in Luganda-English machine translation. In *Proceedings of the 15th biennial conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*. 274–286.

[90] Alicia Beckford Wassink, Cady Gansen, and Isabel Bartholomew. 2022. Uneven success: automatic speech recognition and ethnicity-related dialects. *Speech Communication* 140 (2022), 50–70. doi:10.1016/j.specom.2022.03.009

[91] Genta Indra Winata, Samuel Cahyawijaya, Zihan Liu, Zhaojiang Lin, Andrea Madotto, Peng Xu, and Pascale Fung. 2020. Learning Fast Adaptation on Cross-Accented Speech Recognition. In *Interspeech 2020*. 1276–1280. doi:10.21437/Interspeech.2020-45

[92] S.R. Mahadeva Prasanna Y. Sunil and Rohit Sinha. 2016. Children's Speech Recognition Under Mismatched Condition: A Review. *IETE Journal of Education* 57, 2 (2016), 96–108. arXiv:https://doi.org/10.1080/09747338.2016.1201014 doi:10.1080/09747338.2016.1201014

[93] Sanghyun Yoo, Inchul Song, and Yoshua Bengio. 2019. A highly adaptive acoustic model for accurate multi-dialect speech recognition. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 5716–5720.

[94] Eunseop Yoon, Hee Suk Yoon, John Harvill, Mark Hasegawa-Johnson, and Chang Yoo. 2023. INTapt: Information-Theoretic Adversarial Prompt Tuning for Enhanced Non-Native Speech Recognition. In *Findings of the Association for Computational Linguistics: ACL 2023*, Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (Eds.). Association for Computational Linguistics, Toronto, Canada, 9893–9902. doi:10.18653/v1/2023.findings-acl.627

[95] Emre Yılmaz, Astik Biswas, Ewald van der Westhuizen, Febe de Wet, and Thomas Niesler. 2018. Building a Unified Code-Switching ASR System for South African Languages. In *Interspeech 2018*. 1923–1927. doi:10.21437/Interspeech.2018-1966

[96] Yuanyuan Zhang, Aaricia Herygers, Tanvina Patel, Zhengjun Yue, and Odette Scharenborg. 2023. Exploring data augmentation in bias mitigation against non-native-accented speech. In *2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*. IEEE, 1–8.

[97] Yuanyuan Zhang, Yixuan Zhang, Bence Halpern, Tanvina Patel, and Odette Scharenborg. 2022. Mitigating bias against non-native accents. In *Interspeech 2022*. 3168–3172. doi:10.21437/Interspeech.2022-836

[98] Yuanyuan Zhang, Yixuan Zhang, Bence Mark Halpern, Tanvina Patel, and Odette Scharenborg. 2022. Mitigating bias against non-native accents.. In *Interspeech*. 3168–3172.

[99] Yixuan Zhang, Yuanyuan Zhang, Tanvina Patel, and Odette Scharenborg. 2022. Comparing data augmentation and training techniques to reduce bias against non-native accents in hybrid speech recognition systems. In *1st Workshop on Speech for Social Good (S4SG)*. 15–19. doi:10.21437/S4SG.2022-4

[100] Zhikai Zhou, Wei Wang, Wangyou Zhang, and Yanmin Qian. 2022. Exploring Effective Data Utilization for Low-Resource Speech Recognition. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 8192–8196.

[101] Juan Zuluaga-Gomez, Amrutha Prasad, Iuliia Nigmatulina, Seyyed Saeed Sarfjoo, Petr Motlicek, Matthias Kleinert, Hartmut Helmke, Oliver Ohneiser, and Qingran Zhan. 2023. How does pre-trained wav2vec 2.0 perform on domain-shifted asr? an extensive benchmark on air traffic control communications. In *2022 IEEE spoken language technology workshop (SLT)*. IEEE, 205–212.

Received 24 March 2025; revised 4 September 2025; accepted 15 September 2025