Auditing Algorithmic Interviewers in Ghana: A CBEAT Framework Analysis of Voice-Based AI Hiring Bias and the Imperative for Emotional Data Governance

Oppong Stephen A [1*], Acheampong William [2], Beverla Edwina [3], Beverla Edna [4], Acheampong Akwasi [5], Yuornuo Theodora [6], Benyin Chris [7]

[1] Department of Psychology and Marketing and Entrepreneurship, University of Ghana, Legon, Accra, Ghana
[2] Department of Geography and Environmental Studies, New Mexico State University, Las Cruces, NM 88001, USA
[3] College of Humanities, University of Ghana, LG 25, Legon, Accra, Ghana
[4] Information Services Department, University of Ghana, LG 25, Legon, Accra, Ghana
[5] Baikal School of Brics, Irkutsk National Research Technical University, Irkutsk, 664074, Russia
[6] Department of Ghanaian Languages and Linguistics, University of Cape Coast (UCC), 233, Ghana
[7] Faculty of Computing and Information Systems (FoCIS), Ghana Communication Technology University, 00233, Ghana

* Corresponding Author: **Oppong Stephen A**

Article Info

P-ISSN: 3051-3383
E-ISSN: 3051-3391
Volume: 07
Issue: 02
Received: 12-05-2026
Accepted: 10-06-2026
Published: 08-07-2026
Page No: 36-48

Abstract

In this study we provide the first documented application of the CBEAT Audit framework as an empirical means of evaluating the potential of using voice technologies to assess applicants for employment in the Ghanaian context. Our research assumes that applicant scoring, vocal characteristic recording, and voice transcription represent a set of culturally specific values about emotional expression and communication competence that disadvantage those who speak Fante from Ghana. Our method for analyzing this issue is through an evaluative qualitative single-case study design using only secondary document data as sources. The process consists of five stages involving reviewing documents selected by their level of relevance to the purpose of this study, extracting pertinent details from these documents through an evidence extraction and mapping process, utilizing the evidence gathered through the extraction and mapping stages to create a composite profile of all evidence collected for each dimension of the CBEAT Model (i.e., social and environmental), conducting validation of the findings against three external regulatory standards (i.e., E.U. AI Regulation and the United Nations Convention on the Rights of Persons with Disabilities), and finally mapping all findings onto the institutional context of Ghana. Our results highlight that for each dimension of the CBEAT structure, there exists a combination of aspects that when combined, create a cumulative risk profile that would be impossible to detect if using only one metric per dimension. For example, the word error rates of automatic speech recognition error systems compared to standard test weights for Ghanaian Akan (Fante) speech provide an 80.3% error rate on commercial systems as opposed to approximately a 5% rate on standard test weights. The issue of regulatory restriction on inference of emotionality from speech scoring systems was also explored. For instance, in the case of the E.U. AI Act, the inference of emotionality for workplace purposes is prohibited under Article 5(1)(f). However, the Ghana Data Protection Act of 2012 does not define biometric information and therefore does not prohibit using emotional inference to score an applicant's suitability for a position in the workforce. The overall contribution of this study is three-fold: 1) it will enable a new and practical method of replicating the implementation of a previously conceptualized audit framework (CBEAT); 2) it provides a unique cumulative risk assessment based upon multiple sources of relevant data; and 3) it will provide the Ghanaian government with sound and effective policy recommendations regarding how to govern emotional data in Ghana. This research is timely as Ghana moves towards establishing a national regulatory framework for artificial intelligence through the drafting of a new Data Protection Bill and establishment of a Responsible AI Office; thus, this audit will serve as a significant input for policy development by regulators who will be responsible for crafting legally enforceable standards to govern the use of AI within Ghana.

DOI: https://doi.org/10.54660/IJAIET.2026.7.2.36-48

Keywords: Cultural Bias Evaluation for Affective Technologies, Voice-Based Artificial Intelligence Hiring Assessment Systems, Emotional Data Governance in Ghana, AI Governance Infrastructure, Vocal Characteristics in Hiring Assessment, Data Protection Bill in Ghana

1. Introduction

As artificial intelligence (AI) continues to rapidly expand in human resource management, there is now a new class of algorithmic decision support devices that transform the evaluation and selection of job candidates by human resources professionals. There has been much focus on voice-based AI hiring assessment algorithms that utilize voice data collected from the job candidate while conducting a job interview to record, transcribe, and analyze the unique attributes of the candidate's voice and generate scores or rankings that predict adequate future levels of success at work, personality profile, and emotion status [1]. The vendor community proponents validate their claims that voice-based AI hiring assessment algorithms decrease human bias, enhance efficiency, and allow for the screening of a large volume of job applicants in a scaled manner as a reason to recommend using voice-based AI hiring assessment algorithms [2]. Nonetheless, an emerging body of research is now beginning to articulate how the AI systems have been utilized to reinforce and exaggerate social disparities as they relate to race, sex, accent, and nationality [3]. The overall segment of the market focusing on affective computing technologies (such as voice-based emotion recognition) is expected to experience exponential growth, reaching an estimated $30 to $50 billion multiple for the segment by the end of this decade. There are currently multinational companies that perform international operations in Africa, and a significant number of these companies have large operations based in Ghana and have placed pilot programs or actual use of these systems for graduate recruitment and entry-level hiring [3]. To now use such a tool in a locale that is so different from the culture and language of people in a country where the tool(s) have been designed will raise a number of questions regarding the appropriateness of using these tools in Ghanaian society.

Given the high levels of participation in the workforce of young people in Ghana, the great overall number of Ghanaians that are seeking employment in the form of graduate employment in multinational companies, and the fact that the Ghanaian language is primarily composed of tonal first languages (e.g., Akan, Ewe, and Ga) and their respective tonal elements influence the way that a speaker communicates, Ghana will serve as a particularly germane setting in which to investigate the cross-cultural appropriateness of using voice-based AI hiring assessment [4]. The primary hypothesis associated with this investigation is based on the view that voice-based hiring assessment algorithms have embedded cultural biases in the description of emotional expression, as well as on the view that the systems in which these algorithms operate have built into them an ingrained systemic disadvantage to the use of Ghanaians as speakers. There is substantial evidence from previous researchers with respect to how voice-based hiring assessment (and other uses of voice-based AI) serve to encode systematic discrimination against speakers based on the way that the training corpus of voice samples has been constructed based on individuals from North America and Europe, who continued to speak the English language with a distinctive tonal first language [5]. Once a Ghanaian candidate has been evaluated and receives a score (or rank) based on a voice-based AI hiring assessment algorithm, the resulting scores that are assigned to candidates will produce a series of errors that range from their not being correctly transcribed and/or scored to the exclusion of the Ghanaian candidate entirely and to a major level of inaccuracy in the evaluation process, which a single fairness measure will most certainly not capture.

For the purpose of this research, the development of a systematic audit process for implementing the Cultural Bias Evaluation for Affective Technologies (CBEAT) model as it relates specifically to using voice-based emission AI hiring assessment algorithms will be operationalized through the establishment of the approach to the assessment of voice-based AI hiring assessment algorithms either used or to be used in Ghana. Following the completion of this systematic audit of the selected voice-based AI hiring assessment algorithm(s) and subsequently using the findings from the audit of the selected algorithm(s) to cross-reference with three mature examples of regulatory efforts: The AI Act of the European Union, the Illinois Artificial Intelligence-Video Interview Act, and New York City Local Law #144 (NYC-LL144). After this has been conducted and all findings that are confirmed by the audits and through cross-validation are mapped to the institutional environment in Ghana that governs the collection and/or use of emotional data, including the Data Protection Act 2012, the Data Protection Commission, the forthcoming Data Protection Bill, and the Responsible AI Office, policy recommendations will be provided regarding the governance framework for emotional data collection/use [6]. There are three major contributions associated with this research project. First, this is the first time that the CBEAT framework has been employed in a manner such that it provides an empirical basis for use of the framework beyond being a theoretical concept as a structured methodology and readily replicable in other contexts/technologies beyond this research project. Second, the resulting empirical analysis will provide insight into how the use of voice-based AI hiring algorithms yields a compounded harm profile for the Ghanaian candidates.

In addition to associating a foreign yardstick for cultural evaluation with the chain of bias origins that underpins voice-based hiring algorithms, there are three stacked error types that have been identified for the Ghanaian candidates. Additionally, there are self-reinforcing harm trajectories associated with the Ghanaian candidates, which cumulatively result in systematic disadvantages for Ghanaian candidates, and these disadvantages are sufficiently complex so as to escape the identification of a singular metric that will provide a full snapshot of the issues. Third, based upon the validated results of the audit process, specific policy recommendations for Ghana will be provided with respect to the emotional data governance framework of the country. The timing of these recommendations will align with the ongoing development of the framework for AI governance being established in Ghana through the new Data Protection Bill and the establishment of a Responsible AI Office [7]. The remaining sections provide a general summary of the analysis conducted for the remainder of the paper. The Primary section (Section 2) contains literature that describes current voice-based AI hiring/recruitment methodologies and algorithmic bias as it relates to discussion on the algorithmic bias issues found within the field of affective computing technologies and the Ghanaian AI governance ecosystem. The secondary section (Section 3) provides the theoretical foundation of the CBEAT.

The CBEAT model has been designed as a theoretical diagnostic tool to assist researchers and organizations in evaluating affective technologies employed in the marketplace. The third section (Section 4) provides details about the methods of the research design and the five stages associated with the operationalization strategy for the CBEAT structure. The fourth section (Section 5) presents the results from the audit of the selected CBEAT dimensions of performance as applied to the current phases of the regulatory benchmarking analysis. In section five (Section 6), the discussion includes the implications for the emotional data governance framework within Ghana based on the findings of the research. The recommendations provided for the governance of emotional data [8] will be informed by ongoing efforts to establish national guidelines (e.g. the Data Protection Bill, the Data Protection Commission, and the Responsible AI Office) governing the use of AI in Ghana through the forthcoming Data Protection Bill and the establishment of a Responsible AI Office [9]. Section 7 concludes the paper with an overview of the findings derived from this project and recommends future research designs to provide additional empirical evidence associated with the work described in the paper.

2. Literature Review

The convergence of voice-based AI, hiring algorithms, and cultural bias has become a popular research topic over the past ten years, but there are still many gaps in theoretical frameworks and empirical evidence, particularly with respect to the African context. A large body of research on ASR systems has consistent findings of performance differences between demographic groups. Specifically, Koenecke *et al.* [8] present evidence that commercial ASR systems developed by Amazon, Apple, Google, IBM, and Microsoft demonstrate much higher word error rates for African-American speakers in comparison to white speakers, despite using the same quality audio and accounting for other potential confounding variables. This finding builds on the notion that ASR systems are developed primarily using North American and European English speakers, leading to systematic poor performance for speakers from different varieties of English [9]. Moreover, the issue is not limited to transcription errors but extends beyond ASR to the ability to assess more complex characteristics based on vocal attributes, such as inferring a candidate's emotional state, personality, or job-competent behaviours from voice alone. Recent studies in the area of affective computing have raised concerns about the validity of using vocal expressions to infer these characteristics and have found that many of the most popular commercial systems lack adequate evidence supporting the underlying models of emotional behaviours [1]. Barrett *et al.* [10] argued that the basic model of emotion measurement, which most commercial systems use, does not have adequate empirical support and that therefore assessments of behaviours associated with emotional expression can differ significantly by culture. The relevance of this to the area of voice-based hiring systems is magnified because many commercial systems claim to evaluate characteristics, including enthusiasm, confidence or communication competence, based on prosodic features that are differentially distributed in different languages and cultures [11].

In addition, a separate but related line of research has examined bias in AI hiring systems (e.g., bias in resume screening, video interview evaluations, and personality assessments), with Raghavan *et al.* [12] providing a comprehensive review of commercial AI hiring systems. The review concluded that many vendors provided misleading or unsubstantiated claims about their ability to mitigate bias in hiring algorithms while lacking transparency regarding their data training, model building, and validation processes. In response to these issues, various legislative efforts have been enacted to address the use of AI in hiring, including the Illinois Artificial Intelligence Video Interview Act, which was one of the first pieces of legislation to require that employers inform applicants if their application will be subjected to AI analysis, require consent from candidates, and disclose how AI was being utilized [13]. New York City Local Law 144, which became effective in 2023, further increased accountability by requiring independent audits of automated employment selection tools, although enforcement of this law has been inconsistent to date [14]. The European Union's AI Act is the most comprehensive set of regulations to date regarding AI and categorizes emotion recognition AI systems as "high-risk," while restricting them from being used in workplace and educational settings under Article 5(1)(f), with sanctions for violation potentially reaching €35 million or 7% of global revenues [15]. This classification reflects a growing consensus from international regulatory bodies that there is not enough scientific evidence to support the validity of making high-stakes employment decisions based on emotional inference from vocal or facial expression.

Although research on AI governance and algorithmic bias in Africa is still in its infancy, it is rapidly growing. In 2024, the African Union approved the Continental AI Strategy, which requires African countries collectively to establish national frameworks to govern the ethical, socially inclusive, and human rights-based use of AI in Africa [16]. Ghana is leading the charge in the development of these researches, with the establishment of the Office of Responsible AI in 2024 and with a draft Data Protection Act that is meant to supersede the previous 2012 Data Protection Act [7]. However, there remain significant gaps in the current regulatory framework in Ghana, as noted by Adams & Nkansah [6], with a lack of definitions of biometric data, missing prohibitions on emotion inference, and a loophole in contracts that might make recruitment subject to exemptions from protections implemented in an automated decision-making environment. The CBEAT Framework, developed by [17], was specifically designed to address the aforementioned limitations with existing bias audit frameworks when applied to affective computing. CBEAT consists of five different dimensions of bias that are interconnected to one another, which differ from most existing audits that are only focused on a single dimension of bias (demographic parity vs. equalized odds). CBEAT's five dimensions of bias include: Cultural context, bias origin, error types, affected domain, and trajectory of harm. By identifying the multiple opportunities where bias may enter into a voice-based hiring system pipeline, the CBEAT Framework provides an appropriate way to examine biases during each stage of the hiring process. This study expands upon previous research as follows: previous studies have examined ASR research, emotion-by-voice inference, and AI legislative frameworks individually; however, the current study integrates these three components into a single model to conduct bias auditing of voice-based hiring systems within the Ghanaian context. Although it was originally intended as a theoretical diagnostic tool, this study provides the first empirical (non-theoretical) operationalization of the CBEAT Framework by employing secondary-document analysis to systematically apply the CBEAT Framework to voice-based hiring. Additionally, the focus on Ghana has both timeliness and relevance, as Ghana is currently developing its AI governance infrastructure and thus presents an opportunity to establish an evidence-based regulatory framework that mitigates against the types of harms identified by this analysis.

3. Theoretical Background: The CBEAT Framework and Voice-Based AI Hiring

We utilize the CBEAT framework [17], which was introduced by Hauser and Dignum as a theoretical diagnostic paradigm for describing the existence of bias in affective computing systems, which are defined by their ability to infer, interpret, or react to the human emotional state. The CBEAT scheme enables us to investigate our particular research context because it provides a broader view of the culturally located and multidimensional nature of the treatment of bias in the context of the hiring processes operated on by many AI hiring systems worldwide. The CBEAT system organizes the analysis of bias across five interrelated dimensions. The first dimension, Cultural Context, examines the cultural assumptions embedded in the design of a system, the training data to create a model, and the environment in which the model is applied. For example, the cultural contexts around the design standards for voice-based hiring systems could be examined using the CBEAT solution by examining the cultural bias of choices about what vocal characteristics constitute 'norms' to be used as standard criteria for this system, regarding acceptable and appropriate emotional expression for a standard of professionalism, and about how one's ability to communicate in an assumedly unbiased manner is defined. The second dimension, Bias Origin, explores the various ways that the identification and introduction of bias can occur in the life cycle of the system. For example, there are multiple opportunities for introducing bias into the model and training data used to create an affective technology.

Bias could occur through the selection of an unrepresentative training corpus comprised of examples of non-Ghanaians or the use of culturally specific annotation guidelines for building the training data sets. Additionally, if the deployment practices for the AI hiring system do not account for local language variability or cultural meanings and uses of the aforementioned concepts, this dimension can be used to evaluate that as well. The third dimension, Error Typology, categorizes the errors produced by an affective technology and examines how those errors disproportionately affect demographic groups. The errors of a voice-based hiring system might include producing an incorrect transcription (high word error rate for candidates speaking with a Ghanaian English accent), an incorrect classification (incorrect inference regarding a candidate's shared emotional state and/or personality type), and an incorrect score assigned to the response (underestimating the level of job-necessary skill demonstrated). Within the third dimension, the CBEAT framework defines the target domains impacted by the use of the voice-based AI hiring system (welfare) and provides a definition of statistical disparity as well as a definition of the substantive harm incurred to the individual and the social actor. The fourth dimension, Affected Domain, specifies the areas or aspects of welfare potentially negatively impacted by the voice-based AI hiring system's errors. This domain can also identify the potential negative impacts to the individual as well as their local community. This last point is critical for fully capturing the cumulative effects of bias on individuals and communities.

The fifth dimension, Trajectory of Harm, examines the cumulative and increasing impact of the compounded experiences of bias and its many forms in time across multiple intervention points within the decision-making process. For example, the CBEAT framework may indicate that a voice-based AI hiring system's initial transcription error or biases may create a cycle of similar events occurring across all stages of the decision-making process leading to a candidate's exclusion from the hiring system. This cascading effect of a voice-based AI hiring system has critical implications for understanding why single-point error rate intervention, such as equal error rates during transcription, cannot be used as an effective intervention for addressing the potential for structural exclusions in hiring. In employing the CBEAT framework in the context of voice-based AI hiring systems, it is critical that the specific context and technology being considered include careful operationalization of all dimensions of the framework to capture key features of how bias may have entered the system across the lifecycle of that particular hiring system. For the determination of all three points of entry occurring during the following three stages of voice-based hiring systems: (1) Automatic Speech Recognition (ASR) transcribes voice samples into text information on the recorded response of the candidate to determine the validity of candidate hiring; (2) Natural Language Processing (NLP) analyses and draws conclusions about the quality, relevance, content, and/or sentiment of the text from the ASR stage; and (3) Prosodic Analysis determines vocal features such as pitch, tempo, intonation, and energy to identify evidence of a candidate's emotional response, personality type, and ability to effectively communicate and demonstrate competency for the voice-based hiring systems. Specifically, the ASR stage utilizes acoustic models developed using predominantly North American or British English voices as the training corpus.

When the ASR technology uses the same model to generate transcriptions for Ghanaian English-speaking candidates, bias becomes a factor in the transcription process due to the presence of multiple meanings for similar words in both languages and the application of culturally specific standards for what constitutes 'normal' or acceptable in the expression of emotion vocally. The resultant level of erroneous transcription resulting from this cultural bias can lead to the NLP analysis making an erroneous assumption regarding the quality/type of sentiment detected from the candidate based on errors introduced at the ASR stage. Finally, the data used during the prosodic analysis provides requirements regarding the criteria or assumptions related to pitch variability, indicating confidence in intercultural and national contexts, which can be problematic when the respondents are using a tonal-based language(s), such as Akan. In summary, the use of the CBEAT model provides a systematic and objective methodology for evaluating how and where cultural bias may have influenced the AI hiring system as a measure of the compounding harm around the various types of biases that underpin the performance of the voice-based AI hiring system. The use of this multidimensional evaluation approach is an invaluable tool for informing those seeking to develop regulatory strategies that will address the systemic or root cause of bias rather than creating policies focused only on surface-level (statistical) outcomes associated with bias in the hiring process in order to remove unfair discrimination outcomes in hiring.

4. Research Design and CBEAT Framework Operationalization

The study is based on a qualitative, evaluative method (using secondary data), where we will utilize an established evaluative concept and CBEAT to systematically analyze secondary evidence regarding the use of voice-enabled AI-assisted job hiring assessment in Ghana. This study is conducted without collecting any primary data. The five stages used in our evaluation will be outlined below, with descriptions of the procedures utilized in each stage of analysis.

4.1. Stage 1: Systematic Document Identification and Selection

The first step was to define the categories of the five types of documents (i.e., studies published in peer-reviewed journals that examined ASR disparities; SER bias; algorithmic hiring bias; vendor technical documentation and statements about algorithms explainability; and regulations and guidance from governing bodies in relevant jurisdictions; and audit & civil society reports of relevant commercial AI systems; and reputable newspaper articles and legal analyses of voice-based hiring technology). The inclusion criteria ensured that documents met all three criteria: pertinent to voice/speech assessment; verifiable against a primary (authoritative) source; and empirical data based upon a peer-reviewed publication or governmental body's official documentation. The time frame included documents published between 2019 and 2026 with only older foundational theories that continue to serve as the conceptual basis of the CBEAT model allowed in this review. The exclusion criteria provided that unverifiable second-hand or unverifiable claims would be removed; marketing was to be viewed as fact; and no document could be included based on the claims of new studies.

4.2. Stage 2: Evidence Extraction and Mapping

All retained documents were read and coded utilizing a standardized coding template with one row for each CBEAT dimension: Cultural Context, Bias Origin, Error Typology, Affected Domain, and Trajectory of Harm, which allowed for a systematic comparison of all retained documents against each CBEAT dimension. For each dimension, the coders used the provided coding template to list the specific evidence provided for each type of document, its source citation, and an assessment of its strength. The resulting coding matrix attached each fact documented to each dimension of the coding framework and allowed for comparisons of the findings across all source documents. In addition to the code, the coders used the coding template to capture metadata about each source, including publication type, date of publication, and geographic location, which can be used for later analysis of evidence quality and completeness.

4.3. Stage 3: Structured CBEAT Application

We created a unique example of a voice-enabled AI evaluation of candidates from Ghana using the CBEAT proposed solution based on available data obtained from vendors. The example does not refer to any specific vendor's use of a voice-enabled AI system; instead, it was developed using known practice across multiple vendors due to the absence of public information regarding specific vendor names and implementations. The analysis consisted of evaluating how all CBEAT dimensions could be mapped onto the context of a voice-enabled hiring assessment applied to candidates for jobs from Ghana who speak English. We then examined how each harm experience identified through the audit occurred at each classification, decision, and structural level of the hiring process and whether or not they created cumulative or self-reinforcing impacts on the overall hiring pipeline. By completing this audit following a structured format, we were able to evaluate both independent sources of bias and the subsequent growing impact of those biases cumulatively or multiplying across the entire hiring process.

4.4. Stage 4: Cross-Validation Against Regulatory Standards

In order to determine whether the harms identified by the CBEAT application are identified and under control by established regulators, we compared our findings from the audit against 3 different external regulatory sources: the EU AI Act, Illinois AI Video Interview Act, and NYC Local Law 144. For each of the three regulations, we assessed if the harms specifically identified in our analysis, such as lack of scientific validity for emotion inference, differences in transcription across different accents, or lack of accountability with automated decisions, have been explicitly prohibited, regulated, or not addressed by each of the three regulations. The purpose of comparing the results of our audit against these three regulatory standards was to serve as validation for the results of our audit and provide a gap in governance that could be addressed by the evolving regulatory framework in Ghana.

4.5. Stage 5: Derivation of Policy Recommendations

Lastly, we mapped the validated results against the context of the Ghanaian institutional landscape which includes the Data Protection Act 2012 (Act 843), the Data Protection Commission, the soon to be published (as of the time of publication) Data Protection Bill, and the newly created office that governs AI (Responsible AI Office). In examining each of these institutions, we also determined whether there are current provisions in place to prevent the harms identified in our CBEAT analysis, identify where there is a gap, and what policy interventions may be available to close the gap through the institutions we have mapped. Ultimately our mapping effort has generated actionable recommendations tailored to the Ghanaian legal and institutional environment that can guide the future development of Ghana's AI governance and structure. We are clear about the limitations of our methodology. The current research relies on self-reported disclosure by vendors (which tend to be limited and sometimes overly promotional). The single case examined is a composite example and is not a comprehensive survey of Ghana-based deployments due to the lack of publicly available records about each specific deployment in Ghana. Additionally, we relied on secondary data only and cannot directly measure any of the disparate application impacts on Ghanaians. These limits will be returned to in the conclusion, and we outline a few avenues for future primary research.

5. Audit Results: Bias, Errors, and Regulatory Gaps

The composite case of voice-based AI assessment for the hiring of Ghanaian candidates shows that there are multiple types of harm that can create a significant disadvantage for Ghanaian candidates in the CBEAT service due to using one single metric of fairness for this type of application. The audit results of each of the five framework dimensions will be discussed in the next subsections along the line of: 1) a foreign culture standard/objective; 2) a chain of biases in the origins of the candidate; 3) three levels of error; 4) a selfreinforced path of harm generated. These four factors create the structural disadvantage for the Ghanaian candidate in AI-based voice assessment (hiring).

5.1. Cultural Context

The Cultural Context dimension of the CBEAT framework examines whose voice norms have been incorporated into the design of the hiring system of voice-based AIs treated as neutral criteria for competent communication. Voice-based AI hiring systems have encoded a specific cultural standard of what a competent, confident, and hireable speaker sounds like based on the voice data for the population(s), from which the AI is trained and validated. These voice data sets that have been created and used to train and validate commercial ASR and SER systems (just as an example) are heavily biased in representation of speakers of North American and European English and misrepresent many people who did not have a native English accent. For instance, in Ghana, English has many different prosody (tone, rhythm, etc.), but it is very different from the training distribution of both ASR and SER systems. Ghanaian, a country in which English, is one of the official languages, has as its first language the Akan (Twi and Fante) and Ewe (and Ga) dialects of the Tonal languages. Tonal languages utilize variations in pitch to convey meaning, which Ghana has transferred from their first language to English. In addition, Ghanaians may frequently use an intonation that has an upwards trajectory to signify a question, which is often the opposite of the use of this intonation in a North American speaker's communication. When a voice-based AI hiring system uses a specific tone of voice or frequency of speech to portray a candidate's confidence, enthusiasm, or competence, they are evaluating the way each feature is used in the specific context of Ghana, rather than attempting to measure its meaning across the board.

In addition to the prosodic features of speech, these systems have implicit cultural assumptions built into them, such as turn-taking, response delay, and conversation structure. Researchers have studied and documented cultures' varying standards of appropriateness for pause duration, duration of overlap, and response time to documentation, for example. A Ghanaian candidate may have responded to an interviewer by pausing for a longer time period as they were considering their answer and/or following the Ghanaian norm of taking proper care in communicating (thoughtfulness) when they spoke. Therefore, when an AI voice-based hiring system uses the speed at which a Ghanaian candidate has responded, compared to North American norms favouring a rapid response, the system will evaluate their action as a weak indicator of "confidence" or "quick thinking," as opposed to using the more culturally appropriate evaluation criteria. Another important implication from the cultural context dimension is the assumption that these AI voice-based hiring systems do not reflect cultural diversity and instead impose upon the users of the AI voice-based hiring system a norm of communication standard but do not consider that both the cultural model and norm are based upon a different standard/criteria; therefore, an AI voice-based hiring system will establish the parameters of how employees demonstrate good communication.

Therefore, the AI voice-based hiring system gives a specific speech standard to measure against a Ghanaian's speech standard as being better suited to the training based on an association with them as the appropriators of the training distribution, thereby establishing a premise for defining competence in AI hiring systems in the Ghanaian context (as an example). Finally, within the cultural context dimension, there are very few, if any, vendors of voice-based AI hiring systems who conduct local revalidation or adaptation studies upon the introduction of the employees of the new culture or linguistic background. This means that a voice-based AI hiring system will apply the same trained model based on documentations provided by North American and European English speaking countries to the Ghanaian AI voice-based hiring candidate, but the AI voice-based hiring system will not adjust the trained model for local speech patterns, adjust the established model by calibrating the scoring thresholds, and lack the transparency in the assumptions used in creating the AI voice-based hiring system, thus disadvantaging the Ghanaian candidates in their evaluation and ultimately eliminating the opportunity for them to gain employment through the use of AI voice-based hiring systems.

5.2. Bias Origin

The Bias Origins dimension of the CBEAT framework defines and traces the sources of bias throughout the life cycle of a system, recognizing that bias rarely results from one flaw but is instead the result of a chain of related decisions and omissions that build on one another through the different stages of development and deployment. Utilizing the CBEAT framework, we have identified three separate entry points where bias equilibrates in the context of Ghanaian candidates in the voice-based AI hiring systems: each of these represents a different stage of the system's lifecycle. The training data forms the first entry point. The acoustic models that provide the basis for the ASR used in commercial ASR systems are trained on corpora of speech that are systematically biased against candidates with African-accented English. Olatunji *et al*. (18) have shown that the lack of training data that is representative of African accented English is the primary contributor to the poor performance of ASR systems when it comes to recognizing speech that has been produced using an African accent. The ASR system has no knowledge of what Ghanaian English sounds like during training and therefore will fail to transcribe it accurately. This origin of bias in data is not a technical shortcoming that can be easily resolved through fine-tuning; rather, it is a structural decision regarding whose voice is worthy of collection, annotation, and inclusion into a training corpus. Incentives surrounding the collection of training data primarily benefit populations with large, easily accessible populations of anthropological and Western English language speakers living in North America and Europe; thus, African accented speech will continue to be under-represented in data used to train ASR systems. The model architecture/design and the target that is being aimed for when developing a model create our second entry point into developing a bias origin.

Even if the ASR stage were to operate at 100% accuracy, the downstream scoring phase for evaluating candidates includes its own bias origins. For example, if a vendor defines the scoring target as being "close" to the example transcriptions of "high performer" candidates from previous years, the vendor's reference group will include whatever demographics were skewed in the reference group of past high performers. Raghavan *et al*. (12) demonstrate that this mechanism is central to hiring bias; if a group of past high-performing candidates is made up of primarily candidates who are of a given linguistic demographic, then the model will learn to reward vocal patterns associated with that demographic and to penalize other vocal patterns. In the Ghanaian context, if a reference group that is being utilized to train a model for a multinational graduate program has previously drawn upon participant groups that consisted of individuals who had received education via English-only medium schools and were familiar with North American and/or English language norms, the model will systematically score candidates who utilize the standard forms of Ghanaian-accented English less favorably than candidates who have a less favorable form of English (even when both groups possess equal levels of job-relevant competencies). The final entry point into creating bias origins exists within a vendor's implementation of its product. A system that has been validated in one labour market (usually the U.S. or U.K.) will be installed and used in Ghana as-is, without being revalidated, without modifying the scoring thresholds, and without being assessed for prediction validity for Ghanaian candidates.

Deployment origin bias is particularly insidious in that it is not detectable by the operators of the system. Vendors may produce aggregate statistics that indicate that the system performed well on the original validation set; however, due to the deployment of the system without local re-validation, there is no means to detect that the system is scoring Ghanaian candidates erroneously. Moreover, there is no reason for the vendor to provide any incentives towards rectifying this situation. Ultimately, the framework describes how the origin of bias that exists within voice-based AI hiring systems results from a series of connections rather than a singular flaw. The first link connects the absence of representative training data created by not including all African accents in the training data set. The second link connects to a model design that is trained to not take into account the demographics of previous reference groups used for scoring "high performer" candidates. The third link connects to the operational practice of firms in using their systems in a different geographical context without having a separate measurement criterion from that existing within the original test. Each link in the chain of connections may provide an independent risk to compromising the integrity of voice-based AI systems, but collectively, they pose a cumulative risk to compromising the ability to equitably assess an individual within each respective link in the connected chain. Therefore, the chain of origin links connecting the three entities listed above will undermine the use of general fairness metrics for auditing voice-based AI hiring systems. Metrics such as demographic parity will only capture the downstream consequences of bias (find differences in candidate selection rates); they will not necessarily help in identifying the originating source(s) of bias within the various entities of the model.

Alternatively, even metrics such as equalized odds will depend on the definition of an 'equal' group in order to determine if errors have occurred across the various groups based upon the actual outcome, which necessitates a presumption that an accurate outcome can be defined objectively. The use of the CBEAT to identify bias origins relative to the training data, model design and deployment will allow for targeted interventions within the various operational stages of the individual systems. Through experiences engaging in research concerning the potential and current capabilities of voice-based AI technologies, we anticipate that candidates from Ghana will represent individuals being assessed by a model that was never designed to assess them. As previously stated, the training data used in an AI model will not be representative of the voice of a typical Ghanaian; the scoring targets defined within the model have been selected based upon standards that are not reflective of the average Ghanaian's communication norms; and the operational practices utilized when deploying the systems will not account for linguistic variations that exist within the context of the country. Overall, the existence of each of the above-mentioned links creates an unequal opportunity for Ghanaian speakers to be subjected to an accurately assessable determination; therefore, managing a voice-based AI system in order to accommodate each of the above-mentioned components cannot be managed by changing a single performance parameter or threshold.

5.3. Error Typology

The Voice-Enabled Hiring System's Error Typology is not about any one type but describes three different errors made by voice-enabled hiring systems and outlines how these different types of errors will affect Ghanaian candidates differently from one another. Each of these error types has been identified to have a compounding effect such that the total errors that each Ghanaian candidate experiences are the result of three types: (1) a broken input, (2) a fragile inference, and (3) a questionable construct.

Error Type #1 - Transcription Error. Transcription errors occur in the first step of the process for most voice-enabled hiring systems - at the ASR level of the process. Voice-enabled hiring systems rely on commercial systems to transcribe spoken English into text and produce a language model to evaluate the transcribed responses. For that reason, transcription error rates for African accents are significantly higher than for other English accents. Koenecke *et al.* published an average word error rate of 0.35 for Black speakers compared to 0.19 for non-Black speakers across five different voice-enabled hiring systems and an average word error rate of 0.35 for Black speakers compared to 0.19 for non-Black speakers across five different commercial ASR systems [8]. The discrepancy between the results remained statistically significant even after controlling for audio quality and other confounding variables. Olatunji *et al.*'s [18] AfriSpeach-200 study found that pre-trained ASR models produced errors between 3 and 10 times worse for African accents compared to standard benchmark/reference material and that the ASR system produced a word error rate as high as 80.3% on spoken Ghanaian Akan (Fante) English. As a result, when a person who is a Ghanaian candidate responds to an interview question in their native English, they could be inaccurate as high as 80% of the time. Since the scoring components of a voice-enabled hiring system read textual transcriptions produced using ASR models rather than the actual audio recordings themselves, the result of using a transcription that contains numerous inaccuracies to generate scores also results in a score that does not accurately reflect the true ability of a candidate.

Error Type #2 - Inference Error. Inference errors occur in voice-enabled hiring systems when candidates' emotional state, personality traits, or job-related competencies are inferred based on various vocal features (e.g., pitch, tempo, intonation, and energy) even when the transcription of their response was accurate. The scientific basis for inferring emotional states, personality traits, or job competencies from various vocal features is extremely weak. Researchers such as Barrett *et al*. have demonstrated that the model most commercial emotion recognition systems are built on, which is the basic emotion model, has no strong empirical support and that vocal expressions of emotions vary significantly depending upon culture. Although voice-enabled hiring systems state that they can assess a candidate's "enthusiasm," "confidence," or "communication competence" from vocal delivery (e.g., pitch, tempo, intonation, and energy), the actual relationship between vocal features and these inferred traits is culturally distinct. For instance, many Ghanaian speakers speak a tone-based language; thus, pitch variation is a lexical indication rather than an emotional or attitudinal indication. Therefore, when evaluating a candidate's vocal delivery, a voice-enabled hiring system may interpret the rising pitch contour as uncertainty, while in fact this rising tone is just part of that Ghanaian candidate's speaking style.

Error Type #3 - Construct Error. In addition to errors associated with the input and inference phases of the hiring process, construct errors are also present. Construct errors challenge the fundamental assumption that the use of a vocal delivery method to evaluate job performance leads to successful job performance. Even when both the transcription and inference phases are valid, to most people vocal delivery is not necessarily a valid predictor of how an individual will perform in a job. In addition, researchers of employment interviews have long acknowledged that there is a significant positive relationship between structured (i.e., interview guide and job-related skills and knowledge) versus unstructured (i.e., assessment of communication skills and personality) interviews [20]. In addition, voice-enabled hiring systems focus primarily on the delivery of the candidate's voice and do not adequately evaluate the content of the answers to the interview questions. Thus, each one of the three types of errors is also associated with the candidate's culture. Therefore, if voice-enabled hiring systems and their processing characteristics have been developed based on certain speaker types, those systems will give disproportionately low scores to candidates whose vocal delivery has different characteristics than those of the population on which the system was developed (e.g., a Ghanaian candidate who tries to communicate his knowledge and abilities via prosody is considered low confidence and will not be evaluated appropriately).

In summary, the process of developing feedback systems based on the three error types can lead to significant errors when used by candidates from other cultures. Voice-enabled hiring systems process and/or evaluate incorrectly transcribed text, which provides the systems with an incorrect inference of the job-related abilities of a Ghanaian candidate. Additionally, the construct error means that there is little scientific evidence to support the hypothesis that voice delivery can be predictive of job performance. Overall, the cumulative outcome of each error type significantly contributes to the likelihood of a candidate being incorrectly scored. As such, this analysis also indicates that the combination of the three different types of errors makes the likelihood that any one intervention will effectively target the overall bias experienced by Ghanaian candidates. Therefore, interventions must focus on addressing all three types of errors in tandem, since one of the key components of the CBEAT framework, the Error Typology, differentiates that no one-point intervention can fairly address the three types of errors experienced by Ghanaian candidates. For Ghanaian candidates, the reality of the three error types is that they are all placed in a system that is unlikely to provide an appropriate evaluation of their job qualifications based on the information generated by the voice-enabled hiring systems. Specifically, since voice-enabled hiring systems do not appropriately represent what the candidate has said, do not appropriately infer a candidate's personal characteristics based on their vocal delivery, and do not have any empirical support to demonstrate a causal link between vocal delivery and workplace success, the candidate's score produced by the voice-enabled hiring system is not valid. Each of the three error types undermines the validity of the entire evaluation of a candidate's vocal delivery regardless of their significance and demonstrates that there is considerable time-related variability in how a candidate may achieve a given score.

5.4. Affected Domain

The CBEAT framework's Affected Domain dimension describes the specific dimensions of human welfare that are affected by system errors and ensures that the audit captures more than just statistical disparities but substantive harms to individuals and communities. The affected domain of AI-based voice assessments for hiring is employment, specifically that of access to employment, at the applicant screening stage. The applicant screening stage constitutes one of the highest stakes domains of divergent economic structures such as employer systems create for any culture, due to the extent to which the applicant screening system is controlled by an employer gatekeeper. When entire demographic groups are systematically undercut by a hiring automation system at the screening phase of the hiring process, the implications extend beyond the mere production of a historically significant gap between the relative number of qualified applicants; hiring gatekeepers foreclose on the employment opportunities/career potential/economic mobility of qualified individuals and promote systemic, structural inequality from generation to generation. The concentrated effect of hiring automation systems creates the conditions by which North American high volume employers are able to hire large numbers of employees through use of computer-driven applicant filtering systems. Automated applicant screening/hiring systems are typically used by large volume employers to assess thousands of applicants for multiple entry level jobs (e.g., graduate/entry-level positions/customer-facing positions), where applicants receive scores used by employers to determine if they will advance to an interview stage with the employer's human resources or an internal professional recruiter. Automated applicant screening/hiring systems create a high-leverage point by which a systematic vocal penalty on one candidate can result in disproportionate negative impacts for multiple candidates. A candidate mis-scored at the applicant screening stage is eliminated from any further opportunity for employment/career advancement in the hiring process no matter how qualified or capable the candidate may actually be. The availability of youth in Ghana's labour market is becoming acute. Ghana is the second youngest nation globally with an estimated median age of about 21 years; this coupled with an ongoing increase in the number of both graduate/youth applicants from Ghana entering into the Ghana labor market each year, creates a demographic shift in the Ghanaian youth labour force.

There are many multinational graduate schemes available to applicants from Ghana that are also in high demand, offered by companies such as banks, telecommunications, and professional service companies, which make them desirable due to their structured pathways for career development, competitive salaries, and opportunities to achieve senior roles in an organisation. For many young Ghanaians, success at obtaining a graduate position generally represents a critical means of furthering their economic mobility and professional advancement. Given that there are many thousands of applicants for each graduate scheme and only a limited number of graduate slots available, employers are working under optimal conditions for the recruitment of qualified candidates using automated applicant filtering systems. When there are many thousands (specifically tens of thousands or even hundreds of thousands) of applicants vying for placement into a specific graduate programme at a company fabricating hundreds of thousands of applications at a time, human resource departments must find innovative and effective methods of filtering out unwanted or unqualified applicants from the pool of applicants applying for that programme. Automated applicant screening or hiring systems that rely on voice analysis have emerged as a popular choice or solution for large volume employers seeking a way to quickly and objectively identify qualified candidates without having to spend large amounts of time or money on hiring managers conducting comprehensive, in-depth face-to-face interviews with each candidate. The connection between using automated applicant screening/hiring systems and the systematic disadvantage that Ghanaian applicants experience results from the general characteristics of using automated applicant screening/hiring systems that could assist large volume employers in reaching their goal of efficiently getting the right candidate into the right position quickly and at a lower cost than human resources managers utilizing other means/methods of assessing applicants.

However, by utilizing an automated applicant screening/hiring system in a way that systematically disadvantages Ghanaian speakers, the same level of efficiency achieved through using an automated applicant screening/hiring system results in a disservice to many qualified Ghanaian youth. The affected domain dimension also illustrates that the harmful impact of the systematic vocal penalty experienced by an entire cohort of eligible Ghanaian youth is not limited to just the individual candidates who are receiving an unfair assessment of their qualifications. When a system is developed that systematically disadvantages individuals who are Ghanaian citizens and bans their entry into the working world through automated means, the negative impact of this entire scenario does not only affect the individuals who are excluded from entering the job market or who have been assessed negatively as a result of an automated applicant screening system but also impacts the families, communities, and overall economy of the entire group of individuals in this cohort. While some of the individuals who are ultimately eliminated from consideration through an automated applicant screening system may be more qualified or equally qualified than some of those who were placed into an interview stage, they are denied the opportunity to showcase their abilities and skills in the more traditional manner of a human resources manager conducting a thorough, face-to-face interview. The systematic denial of access to employment for an entire cohort of eligible, qualified youth in Ghana will create structural barriers for that population from being afforded the opportunity to access professional employment/careers, leadership roles and economic resources that facilitate access to life opportunities. The participants also will experience an ongoing cycle of inequality as a consequence of this structural disadvantage and will likely not escape from this cycle in future generations even if those individuals ultimately qualify through an automated screening system in the future.

Additionally, the affected domain dimension of the hiring assessment or selection process enables one to fully understand and realize the significant power imbalance in the relationship between a job candidate applying for a job and the hiring employer through automated applicant filtering systems. As there is currently no mechanism available for a job candidate to learn about or be aware of the existence of an automated applicant filtering system or the process used to determine an individual's score by an automated applicant filtering system, or to know the training data used by an automated applicant filtering system, or to have access to the criteria under which a candidate was assessed by an automated applicant filtering system, it can be concluded that a job candidate applying for a position would have no recourse if they believe they have been treated unfairly or if they are not satisfied with their assessment. Moreover, similar to the two states of Illinois and New York providing responses to the concerns and suggestions put forth in relation to these matters through their respective enactments of provisions for obtaining consent and conducting bias audits, there is also no similar set of standards of conduct or parameters to protect against this violation in Ghana. The Affected Domain dimension of the CBEAT framework establishes the basis upon which to ascertain that all of the identified dimensions of the CBEAT model's Potential Areas of Harm–Cultural Context, Source of Bias, and Type of Error–and have the potential for creating actual, real-world harm to Ghanaian candidates with respect to their access to employment using an automated applicant filtering system. The Affected Domain is Employment; the Stage is Applicant Screening; the Context is the Youthful Labour Market and High Demand for Graduate Job Opportunities. Collectively, these factors create an acute sensitivity in the affected domain, as a systematic vocal penalty at this stage of the hiring process has the potential to limit the job opportunities for an entire cohort of qualified Ghanaian youth. The Affected Domain dimension serves to frame the overall goal of the audit effort by relating the results of the technical analysis related to the CBEAT framework–bias and error—from a technical perspective to the tangible impacts that are associated with that technology for individuals and communities.

5.5. Trajectory of Harm

The "Trajectory of Harm" part of the "CBEAT" model is unique, as it takes into consideration how much bias intensifies and changes during different phases of a person-focused technology-based hiring decision-making. In this case study of Ghanian individuals, they are exposed to the compounding trajectory of "Harm" that develops during three stages: the Classification Stage, the Decision Stage, and the Structural Stage, where the contribution of harm increases in magnitude because of the influence of the prior stage. The first stage of the trajectory, classification, is primarily based on an automated system misinterpreting and incorrectly scoring a Ghanian's spoken response to their question about their capabilities. Because most voice-based HR systems today utilize ASR technologies that have been trained on typical English language participants (non-humans), they have little to no capacity or experience assessing a speaker's qualities based on an understanding of the sounds (speech) produced by individuals who were raised in a language using tonal features. As a result, Ghanian individuals will have an artificially low competency signal that reflects a Ghanian's pronunciation (i.e., accent) as opposed to their actual competence or ability to perform the function for which they were eligible [18]. Through the ongoing consequences of the Filter Stages, having been previously misclassified, a Ghanian with the same relevant qualifications and demonstrable skills relative to an individual with a non-Ghanian pronunciation represents a lower classification rank; thus, completing a deletion from consideration of Ghanian candidates to receive recognition for the qualification credentials and relevant capabilities. After filtering through the process of decisions on participation in the current selection (i.e., you were classified below the required threshold for scoring purposes), the Ghanian subsequently becomes eliminated from the process without a human reviewer having reviewed the automatically created scores from the previously established filters nor having given the candidate the opportunity to communicate with the hiring entity (e.g., corporate parent organizations) to participate after obtaining a rejection.

Whether or not a Ghanian is eligible for employment in the context of hiring will continue to be denied because of the misclassification created at the classification stage that led to that candidate falling below an acceptable threshold score in a dated reference frame (i.e., before the time allocated to the current selection process). Likewise, in FBC cases, the results of the system demonstrating that a candidate has higher (or lower) vocal-type characteristics than a candidate who has been previously selected as an FBC failed to provide a pathway for successful Ghanian candidates to be connected to E3(3) partners for consideration. The final stage of the Ghanian trajectory structurally disallows Ghanian candidates from being able to perform at a competitive level, as being part of the same cohort in future employments presents Ghanian input to the available capacity and eliminates all potential opportunity for Ghanian candidates to be included among the high-performing competitors in the future. Consequently, based on the vendor employing the same standardized parameter with regard to the evaluation of the same characteristics of each Ghanian candidate in different hiring cycles, all accumulated misunderstandings continue to influence and direct the Ghanian community population characteristics. Thus, Ghanian candidates are continually eliminated from the equation, making the Ghanian population equally disadvantaged as has happened historically throughout multiple cycles.

In conclusion, the "CBEAT" model reveals the fact that the harm being inflicted upon individuals as a result of employment assessments will repeat itself and not solely occur once but accumulate throughout the trajectory of the duration of time spent as it continues in the hiring process and will account for a large number of individuals who were overlooked, excluded, and underutilized for basic employment opportunities, thereby resulting in an unintentional reproduction of the harm that is being done. The "Trajectory of Harm" model quantifies the potential impacts of automatic filters between Ghanian participants who have entered the job market, as having created non-competitive credentials amongst them, thus producing new applicants who would otherwise have had access to equal opportunities relative to all Ghanian candidates (competitors). Therefore, within the context of the legacy of systemic inequities, this dynamic nature of algorithmic bias has significant implications for the understanding of employment discrimination challenges faced by Ghanian individuals. The "Trajectory of Harm" has significant economic and social implications relative to the benefits provided to Ghanian males and females under the CBEAT, that those who were selected to fill the 3, 5, or 6 employment positions organizationally developed independently (i.e., established) from Ghanian males or females; therefore, they will incur costs when compared to those previously considered to fill those employment positions having been developed within the Ghanian population, under the pre-fashioned entrepreneurial opinion.

6. Implications for Emotional Data Governance in Ghana

This audit's findings suggest considerable consequences for emotional data governance in Ghana, as the latter country moves forward with creating its national AI infrastructure, using the new National Data Protection Bill and the recently established Responsible AI Office. The CBEAT Framework shows that harm caused by AI voice-based hiring systems can be understood through the lens of emotional data governance; however, Ghana currently has no regulations addressing this harm. The Data Protection Act 2012 [21] fails to define biometric data, prohibits inferring emotions, and contains a contract exception that may allow recruitment to be exempt from automated decision protection. This regulatory failure is not simply an oversight but is a key vulnerability that would allow vendors to place systems into operation in Ghana that would be banned in other jurisdictions. Article 5(1)(f) of the European Union AI Act [22] serves as an excellent example of what a robust emotional data governance framework would look like. Article 5(1)(f) prohibits inferring emotions through an AI system when used to affect employment or education, because both emotion inference and AI systems are found to be both unscientific and controversial with regard to their use to determine a person's access to employment or education. Furthermore, the EU recognizes that emotional data is of a different nature than other personal data since it is subjective, culturally relative, and has the potential to be used inappropriately for high-stakes decisions.

Ghana's new Data Protection Bill should consider including an explicit prohibition on the engaging of emotion inference in employment or, at a minimum, require that any AI system that claims to measure emotional state or personality traits from vocal characteristics must demonstrate scientific validity in relation to the specific population to which it is applied. The Illinois Artificial Intelligence Video Interview Act [23] and New York City Local Law 144 [24] each present two options that can be used as a reference to create regulations for Ghana. The Illinois AI Video Interview Act [23] mandates that an employer must disclose to an applicant that they are being evaluated by an AI system. The Illinois law adds that the applicant must provide informed consent prior to evaluating the individual and receive a description of how the AI system functions. New York City's Local Law 144 requires independent bias audits of automated employment decision tools. In Law 144, public disclosure of the audit results is mandated. Nonetheless, New York City's experience demonstrates that simply crafting a well-designed regulation does not guarantee that the regulation will be honored, as evidenced by only one of 32 companies identified with a violation reviewed for enforcement against.

In order to prevent enforcement gaps like those in the New York City Local Law 144, Ghana's Responsible AI Office should establish oversight mechanisms that have adequate resources, technical expertise, and enough authority to enforce the regulatory mandate to ensure that the regulatory requirement translates to candidate protection. Policymakers in Ghana should consider three primary implications of this study. First, the new Data Protection Bill should include a definition of biometric data that encompasses vocal characteristics and physiological signals that can be used to infer emotion in voice-based hiring systems, thereby bringing them within the purview of data protection laws. Second, the new Data Protection Bill should prohibit or impose heavy restrictions on the use of emotion inference in the employment context and serve to close the existing contract exception for automated decision protection in the recruitment process. Third, the Responsible AI Office should create sector-specific guidelines for using AI in hiring, including local revalidation requirements, demonstrating clarity regarding system capabilities and the system's limitations, and providing candidates with means for redress. These recommendations are based on the specific harms uncovered through the CBEAT audit and are intended to translate into actions that would be foundational to creating an environment conducive to safe and effective use of AI in hiring in Ghana.

While Ghana is the specific country addressed in this audit, the implications of this audit extend beyond to algorithmic fairness and governance in general. The CBEAT provides a framework to help researchers identify how to use audit processes to assess the fairness of voice-based AI systems, moving beyond simply utilizing statistical parity measures to consider cultural bias inside the design process, track bias down a chain of bias generation during the life cycle of the AI voice assistant, address the stacking of different types of error, and analyze the time period over which harm occurred. This multi-faceted approach allows an understanding of the disadvantage experienced by speakers of non-standard English varieties due to the continued reliance upon the normative assumption of a culturally imposed vocal norm being considered the accepted standard for all competent forms of spoken communication. The Framework reveals the need for a regulatory benchmark model to identify governance gaps through a comparative analysis of Ghana's regulatory framework to that of the European Union, thereby highlighting additional types of evidence required to support regulatory compliance. The goal of this study is to identify how regulations can close the regulatory gap that exists in the areas of AI and hiring through a comparative analysis of the AI hiring regulations of three jurisdictions–Illinois, Illinois, and New York City. The first step to conducting a successful AI hiring regulations gap analysis is identifying the rationale and logic of AI hiring regulations; therefore, the focus here will be on identifying the impact of the AI regulations on the hiring experience of AI job candidates and employers (the participants in AI hiring) as well as the methodology used to assess the AI hiring regulations.

6.1. Limitations of this Study

As noted throughout this study, this research is limited by its reliance on secondary document analysis only. There was no primary data collection of interviews conducted with either Ghanaian job candidates who have been evaluated by voice-based AI hiring systems, nor were any of the other stakeholders involved in this area of study (Ghanaian AI hiring employers and their vendors) interviewed. Although this research was based on documented evidence from secondary sources, the findings presented herein are hypothetical constructs and do not account for the full complexity of the deployment contexts of AI hiring systems identified by the CBEAT. In addition, while the benchmarks utilized as part of this regulatory gap analysis study were based only on that which existed within three specific jurisdictions–Ghana, Canada, and Brazil–further examination of other countries' regulatory models, such as Canada's Directive on Automated Decision-Making and/or Brazil's Lei Geral de Proteção de Dados, for example, may yield more regulatory models/cases that would provide insights into potential solutions applicable to Ghana. Lastly, the CBEAT system is a diagnostic tool that identifies potential harms that may exist within a given technology, and thus it does not provide technical or policy intervention recommendations; consequently, it requires additional time and effort by Ghanaian stakeholders to translate any audit findings from this study into actionable recommendations to eliminate the identified potential losses that each stakeholder experienced.

6.2. Future Directions for Research:

Based on the results presented in this study, future directions for research include a) further empirical research that captures the real-world experience of Ghanaian job candidates who were evaluated by a voice-based AI hiring system, including their level of knowledge regarding this technology, their individual perception of fairness within this technology, and their experience in the hiring process as job candidates, and b) research examining whether bias patterns documented for the Ghanaian English accents are equally applicable to the accents of Nigerian English, Kenyan English, South African English, etc. (the other varieties of English found in Africa) that have varying prosodic features (i.e., pitch/tone of voice); this type of study will demonstrate whether the cultural myopic perspective that is illustrated in this study can be determined to be a universal characteristic amongst all African accent speakers or whether its extent varies by region, language, etc., and c) development of culturally relevant alternatives to the current methods of voice hiring system assessments needs to occur in order to ensure equitable access to the technology by all job candidates if the foundational hypothesis that vocal delivery is a predictor of job success is to be questioned or rejected; therefore, avenues for exploration may include structured written assessment(s), testing of various technical skills required for the position being applied for, and/or conducting portfolio reviews of the job candidate while implementing an alternative chosen from the listed methods for evaluating job candidates with non-standard speaking styles (i.e., non-standard speakers).

Additionally, further, AI hiring systems developed for the provision of both culturally specific and culturally adaptive features may warrant requiring additional scrutiny as to their reliability; i.e., strong evidence would need to exist within the framework that no systemic bias exists between processing data for candidates with non-standard accents and providing false negative assessment results within any developed solutions. However, caution must be exercised regarding relying solely on technical fixes in order to prevent potential masking of the deeper underlying construct issues associated with why and how speech is connected to job performance. In general, emotional data governance represents an entire area of study that will be future-oriented because much of the existing research within the context of AI governance is focused on the developed areas of Europe, North America, and East Asia; thus, fostering an awareness within the academic communities outside of Africa, the lack of existing empirical evidence regarding emotional data governance or how to develop an emotionally safe AI hiring system for job candidates exists in the developing countries. Ghana's experience with the Data Protection Bill and Responsible AI Office stands as an exemplary model that exemplifies how an emergent country can create effective AI governance frameworks by developing/elaborating regulatory frameworks for the deployment of AI in hiring that represent both culturally specific and internationally shared values.

Further, the comparative research conducted across the African countries that are at different stages of AI regulation will identify areas of commonality and sufficient variation across a number of the countries for the purpose of informing both the processes and outcomes of building effective emotional data governance frameworks. Furthermore, various other research needs exist with respect to the political/economic aspects associated with the AI deployment within African labour markets, such as identifying the role of various multinational vendors involved, establishing firm and governmental incentives for employers to implement successful AI hiring systems, and developing an understanding of the power dynamics that shape the development of AI regulatory frameworks. Ultimately these areas of research will provide the cultural context necessary to develop both technically sound and politically viable governance interventions.

7. Conclusion

As a result of this study's application of the CBEAT framework to the Ghanaian context, we were able to identify three types of errors (transcription, inference, and construct) that together create systematic disadvantages for candidates from Ghana when using AI hiring assessment tools powered by voice technology. With the introduction of our regulatory benchmarking exercise, it was also discovered that the country's Data Protection Act 2012 does not address any of the harms identified in this research, while the European Union, Illinois, and New York City have enacted legal guidelines that could serve as a model for Ghana's future Data Protection Bill and Responsible AI Office. The three major contributions of this research are as follows: (a) We redefined an existing theoretical framework and developed an audit methodology that can be utilized for research studies utilizing other forms of hiring assessment and other geographies; (b) We developed a multi-dimensional view of the harm profile with a focus on the compounded effects of the cultural context, where the bias originated, the type of errors that occurred, what domains were affected, and the cumulative harm trajectory over time; and (c) We recommended specific policies based on evidence of the specific vulnerable population of Ghana's youth labor market and the emerging nature of the country's AI governance system. Future research directions should include additional primary empirical studies focused on documenting the actual lived experiences of Ghanaian job candidates who have been exposed to AI hiring assessment tools through voice technology; designing comparative studies between various African English dialects to determine whether the same patterns of bias exist in those dialects; and developing non-penalizing culturally appropriate alternatives to voice-based job assessments. The lack of scholarship examining the emotional data governance gap within Africa is an urgent need for further investigation, as there continues to be a need for multinational companies seeking to expand their deployment of affective technologies into regions with limited regulatory plans. The findings of this research provide strong support for CBEAT as a valid solution for measuring the full extent of harm faced by speakers of non-standard English dialects and providing actionable recommendations for developing culturally relevant and internationally compatible governance interventions for this population.

References

1. Oppong S. Emotion recognition artificial intelligence and the psychology of human affect: a critical examination of theory. ResearchGate Preprints. 2026. doi:10.13140/RG.2.2.24716.32641.
2. Anandhi T. Ethics and governance in AI: bias, fairness and accountability. In: Artificial Intelligence: Foundations, Frontiers and Applications. 1st ed. Boca Raton (FL): CRC Press; 2025. Chapter 4. doi:10.1201/9781003460831.
3. Tembine H, Tapo A, Danioko S, Traoré A. Machine intelligence in Africa: a survey. Authorea Preprints. 2024. doi:10.22541/au.170664213.91680191/v1.
4. Kwarkye TG. Ghana's pathway to AI governance and its implications for Africa. Institute for Security Studies Papers. 2025;418:1-24. Available from: https://hdl.handle.net/20.500.12592/9n4mj35
5. Tegon R. Toward the 4th Agenda 2030 goal: AI support to executive functions for inclusion. In: Handbook of Research on Teaching with Virtual Environments and AI. IGI Global; 2021. p. 204-228. doi:10.4018/978-1-7998-7638-0.ch010.
6. Asamoah JK, Bewel DB. Reimagining Ghana's Electronic Transactions Act in the age of artificial intelligence: legal gaps, risks, and opportunities. J Law Glob Policy. 2025;10(3):128-149. doi:10.56201/JLGP.vol.10.no3.2025.pg128.149.
7. Okolo C, Aruleba K, Obaido G. Responsible AI in Africa: challenges and opportunities. In: Social and Cultural Studies of Robots and AI. Cham: Springer; 2023. p. 115-142. doi:10.1007/978-3-031-27470-1_6.
8. Jahan M, Sun Y, Mazumdar P, Fagyal Z, *et al*. FaiST: a benchmark dataset for fairness in speech technology. In: Proceedings of INTERSPEECH 2025. 2025. p. 1421-1425. doi:10.21437/Interspeech.2025-1892.
9. Ngueajio M, Washington G. Hey ASR system! Why aren't you more inclusive? Automatic speech recognition systems' bias and proposed bias mitigation techniques: a literature review. In: Proceedings of the International Conference on Human-Computer Interaction (HCII 2022). Lecture Notes in Computer Science. Vol. 13303. Cham: Springer; 2022. p. 523-540. doi:10.1007/978-3-031-05412-9_38.
10. Tcherkassof A, Dupré D. The emotion-facial expression link: evidence from human and automatic expression recognition. Psychol Res. 2021;85(8):2953-2969. doi:10.1007/s00426-020-01448-6.

11. Akçay MB, Oğuz K. Speech emotion recognition: emotional models, databases, features, preprocessing methods, supporting architectures, and challenges. Speech Commun. 2020;116:56-86. doi:10.1016/j.specom.2019.12.001.
12. Raghavan M, Barocas S, Kleinberg J, Levy K. Mitigating bias in algorithmic hiring: evaluating claims and practices. In: Proceedings of the 2020 ACM Conference on Fairness, Accountability, and Transparency (FAccT '20). 2020. p. 469-481. doi:10.1145/3351095.3372828.
13. Illinois General Assembly. Artificial Intelligence Video Interview Act. 820 ILCS 42. 2020. Available from: https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015
14. Jansen L. The employment effects of AI hiring transparency: evidence from New York City's Local Law 144. SSRN Electronic Journal. 2026. SSRN:6465063. doi:10.2139/ssrn.6465063.
15. Musch S, Borrelli M, Kerrigan C. The EU AI Act: a comprehensive regulatory framework for ethical AI development. SSRN Electronic Journal. 2023. SSRN:4549248. doi:10.2139/ssrn.4549248.
16. Ajuzieogu U. The African Union AI continental strategy and the development-governance paradox. ResearchGate Preprints. 2025. doi:10.13140/RG.2.2.18932.77440.
17. Barker D, Tippireddy MKR, Farhan A, Ahmed B. Ethical considerations in emotion recognition research. Psychol Int. 2025;7(2):43:102-119. doi:10.3390/psycholint7020043.
18. Elmadany AA, Kwon SY, Toyin HO, Inciarte AA, Aldarmaki H, Abdul-Mageed M. Voice of a continent: mapping Africa's speech technology frontier. In: Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2025. p. 3102-3120. doi:10.48550/arXiv.2505.18436.
19. Lin Y, Tsai Y, Chen K, Huang H, Chou H, *et al*. Toward fair speech technologies: a comprehensive survey of bias and fairness in speech AI. arXiv [Preprint]. 2026. arXiv:2605.01597. doi:10.48550/arXiv.2605.01597.
20. Oswald FL, Behrend TS, Bhatia AA, Morelli KL. Artificial intelligence and employment assessment: future directions for science, practice, and policy. Annu Rev Organ Psychol Organ Behav. 2024;11:411-437. doi:10.1146/annurev-orgpsych-110721-035832.
21. Lynskey O. The foundations of EU data protection law. Oxford: Oxford University Press; 2019. doi:10.1093/acprof:oso/9780198718307.001.0001.
22. European Parliament, Council of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. 2024;L 2024/1689. Available from: http://data.europa.eu/eli/reg/2024/1689/oj
23. İşcen ÖE. The geopolitical aesthetic of computational media: media arts in the Middle East [dissertation]. Durham (NC): Duke University; 2020. Available from: https://hdl.handle.net/10161/21462
24. Carrier R, Brown S, Hickok M, Miller C, McCarthy M, Chung E, *et al*. New York City Local Law 144: bias audits for automated employment decision tools. Center for AI and Digital Policy/ForHumanity Report. 2021. Available from: https://forhumanity.center/web/wp-content/uploads/2023/10/New-York-City-Bias-Audit-An-Overview-and-Action-Plan-v2.pdf
25. Vecchione B, Qu T, Cai P, Smith A, *et al*. Null compliance: NYC Local Law 144 and the challenges of algorithm accountability. In: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24). 2024. p. 2012-2024. doi:10.1145/3630106.3658998.

How to Cite This Article

Oppong SA, Acheampong W, Beverla E, Beverla E, Acheampong A, Yuornuo T, et al. Auditing algorithmic interviewers in Ghana: a CBEAT framework analysis of voice-based AI hiring bias and the imperative for emotional data governance. Int J Artif Intell Eng Transform. 2026;7(2):36-48. doi:10.54660/IJAIET.2026.7.2.36-48.

Creative Commons (CC) License