



American Journal of Applied Research and AI (AJARAI)

VOLUME 1 ISSUE 2 (2026)



PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

The Relationship Between Artificial Intelligence Systems, Voice-Based Technologies and the Sociolinguistic Formation of Hausa People Within the Developing Digital Domain

SHUAIBU Abdulwaheed^{1*}

Article Information

Received: November 09, 2025

Accepted: January 01, 2026

Published: October 13, 2026

Keywords

Artificial Intelligence, Hausa Language, Language Ecology, Sociolinguistic Identity, Voice Technology

ABSTRACT

The fast proliferation of voice technologies based on artificial intelligence prompted the concern of how indigenous languages will survive and have access and visibility in cyberspace. This paper analyses how the Hausa language fits in the changing digital linguistic space, but more specifically, AI-based voice recognition and sociolinguistic consequences. The research is based on the Language Ecology Theory of Haugen (1972), in which Hausa is conceptualised as an interactive ecosystem that is influenced by sociocultural, technological, and environmental factors. The qualitative research design is chosen based on a critical analysis of current AI voice systems, an analysis of voice-enabled apps, and ideas on the subject from appropriate academic sources and opinion leaders. It has been found out that Hausa is still far under-represented in large AI voice technology systems because of inadequate training datasets, orthographic variation, and dialect. The consequences of this exclusion include technological access, linguistic presence and retention of sociolinguistic identity among Hausa speakers. This paper states that language representation of AI systems is a technical and ethical problem, which is inextricably connected with digital rights and cultural sustainability. It is thus supportive of culturally responsive AI design by creating Hausa voice corpora, working with linguists and developers, and community contributions. In general, the paper makes significant contributions to the existing discussions about linguistic equity and inclusive AI development in multilingual communities.

INTRODUCTION

Both artificial intelligence and voice technology modify the way languages function on digital platforms when dealing with the processing and preservation of language materials. These modern communication devices have started to establish deep roots in everyday systems which generate complex effects for the 50 million speakers of Hausa. The prevalence of English and other worldwide languages inside AI systems shows how African languages get excluded from modern digital infrastructure development, according to Osei (2021).

The voice technology space demonstrates discrimination through its service provision because Google Assistant and Siri operate primarily for native speakers of powerful world languages. Eberhard *et al.* (2023) report that Hausa belongs to the top African languages for speakers yet digital access for this language remains limited. The digital gap exists in two dimensions which combine technology with sociolinguistic elements. Makoni and Pennycook (2007) explain that technical systems come with pre-established ideologies that establish linguistic status rankings, which result in marginalisation of languages with limited support.

Elements of sociolinguistic identity for Hausa speakers closely relate to their traditional ways of speaking through tone patterns and formal etiquette rooted in their cultural practices. Most voice models lack the capability to preserve these traits. The research conducted by Birhane (2020) shows how AI systems maintain social discrimination

by processing datasets that remain noninclusive and nondiverse. Voice technology solutions for Hausa may eliminate crucial parts of the language identity because they do not represent the linguistic variations found in the Kano, Katsina, and Sokoto regions.

The risk extends beyond miscommunication. According to Murray (2019), digital tools with standardised language functions cause a subtle process of linguistic invasion by pushing native language distinctions into Western technological patterns. AI systems require placement featuring the Hausa language because this linguistic tradition embraces diverse cultural traditions throughout Nigeria while extending its influence across national and international borders.

The process of integrating Hausa within AI-based voice interfaces goes beyond simple technological progress. The process of "decolonising" language technology requires communities to participate in both data collection and lexicon development and ethical guidelines for language model training, according to Bird (2020). Locally controlled natural language processing programmes together with localised corpus development programmes demonstrate significant potential to translate linguistic goals. Empowering African native languages within contemporary knowledge systems and communication networks remains vital, according to Bamgbose (2011), for stopping global standardisation and protecting cultural independence.

The research examines the connections between artificial

¹ Department of Linguistics and African Languages, Adeyemi Federal University of Education, Ondo, Nigeria

* Corresponding author's email: princedebayo11@gmail.com

intelligence and speech-based systems with linguistic identities through an analysis of Hausa representation in upcoming electronic frameworks. The research intends to strengthen linguistic justice discussions and technological equity debates by recognising Hausa as a Central African language for digital linguistic advancement.

Statement of the Problem

The rapid advancement of modern voice technology and artificial intelligence has led to huge transformations in communication, while the Hausa language, as spoken by millions of West and Central African speakers, remains largely ignored. The common language of Hausa exists outside mainstream voice technology because they mostly serve global languages starting with English. Voice systems present a risk to Hausa cultural protection because they do not support the tonal specificities and dialectal characteristics and sociolinguistic rules of this language.

The absence of the Hausa language from AI programmes restricts its usability while creating an exclusivity problem that marginalises African languages throughout the modern technological realm. The research examines Hausa incorporation into AI systems by studying implementation barriers together with cultural and social implications of leaving this language out of technological devices.

Objectives of the Study

- To analyse how well the Hausa language is included in contemporary AI-powered voice technologies occurs in this research.
- To investigate how to solve integration problems in voice systems for Hausa with emphasis on linguistic characteristics such as tone and dialect differences.
- To investigate both linguistic and social aspects that occur when digital platforms disregard Hausa cultural content.
- To propose methods to improve Hausa AI technology integration together with solutions to preserve and represent Hausa on digital platforms.

Research Questions

- What methods incorporate Hausa speech into artificial intelligence voice technology at present?
- What barriers stand in the way of implementing Hausa tonality together with its dialectal elements into digital speech technologies?
- What social and cultural challenges occur when digital platforms fail to include Hausa as part of their operations?
- Which methods can improve the representation of Hausa in artificial intelligence technologies?

Theoretical Framework

The research will use the Language Ecology Theory of Haugen (1972) as its foundation to understand how languages connect with their surroundings, especially when technology becomes integrated. Language ecology

recognises languages as interactive systems that connect with groups and systems of culture together with technological development. This theory provides the basis to study how digital platforms affect Hausa language vitality as well as how artificial intelligence technologies either aid or restrict its growth. The framework will serve as a guide for Hausa digital space evaluation and establish solutions for building a sustainable digital linguistic framework that includes all languages.

MATERIALS AND METHODS

This research uses a qualitative research design to explore how the Hausa language is positioned in voice technologies that utilise AI. The qualitative research design suits well since the study is exploratory and theoretical, as it is aimed at studying meanings, representations, and systemic challenges, but not numerical measurement.

The literature review is the initial step of the methodology. The literature available on artificial intelligence, voice technology, sociolinguistics and the representation of indigenous languages is reviewed systematically. This covers academic journals, books, policy documents, and technical reports which speak of AI structures and speech recognition systems as well as the online presence of African languages and specifically Hausa. The review assists in defining the knowledge that is available, the gaps in the research and locating the study within current discussions.

The second phase involves the semi-structured interviews. Three groups of participants will be interviewed, including linguists with knowledge of Hausa, AI developers/technologists working with language-related systems, and native Hausa speakers who use digital voice technologies. The respondents will be chosen purposely in terms of relevance and knowledge. The interviews will seek to bring out such insights as data scarcity, dialectal variation, orthographic problems, and the user experience in voice-enabled systems.

Case study analysis of the selected AI voice platforms, including speech recognition tools and translation systems, will be the third stage. These websites are discussed to evaluate the level and quality of support of the Hausa language. Thematic analysis of data of literature, interviews, and case studies identified through the Language Ecology Theory are used to determine the way the technological environments influence the survival and functionality of Hausa in digital spaces.

LITERATURE REVIEW

The development of artificial intelligence (AI) has hugely revolutionised human communication and voice-based devices, including automatic speech recognition and natural language processing. Russell and Norvig (2016) state that AI is conceptualised as systems that can imitate human cognitive processes, such as language comprehension and speech interaction. Along these lines, Jurafsky and Martin (2020) illustrate that modern voice technologies rely strongly on massive linguistic datasets,

which has led to the advantaging of a small number of dominant global languages. Every piece of empirical research in the field of computational linguistics demonstrates that languages that have small digital corpora are still poorly represented in the applications of AI, which supports the existing linguistic hierarchies (Choudhury *et al.*, 2020).

The neutrality of these technologies has been challenged by the critical scholars. Bender *et al.* (2021) state that the language models recreate the socio-political organisations of the information that trained them, thus marginalising the poorly represented linguistic groups. On the same note, Bird (2020) describes the fact that indigenous languages are not used in speech technologies as an extension of colonial dynamics in the digital realm. Although these works are quite useful, as they criticise the current AI trends in the world, they tend to represent the language marginalisation at a more general scale, across all the African languages, without much specificity to particular languages like Hausa.

In the context of African linguistics, Bamgbose (2011) emphasises the long-term marginalisation of the indigenous languages in the formal sphere even though they are actively used. Osei (2021) also applies this argument to the case of AI, stating that African languages receive minimal focus in terms of their appearance in language technology development. The empirical results of the study by Birhane (2020) also indicate that the algorithmic systems have the propensity to increase the inequalities that already exist in the case when the ethical aspect and the situation in the local language are overlooked. Nevertheless, a large part of this research revolves around policy and ethics and the lack of this analytical focus on the sociolinguistic determinants of the influence of AI voice technologies on particular language communities.

Hausa, despite being demographically as well as transnationally dominant, is poorly represented in cyberspace. Jibril (1982) and Malmberg (1981) linguistic research reports the complexity of phonetics and dialects of Hausa, which are technically inconvenient to voice recognition systems. However, as Eberhard *et al.* (2023) reveal, the enormous population of Hausa speakers does not correlate with the lack of digital representation of the language. Although Oluwole (2013) explains that Hausa was particularly dominant in the education and religious sphere, there is not much empirical research done to determine its role in the AI-based voice interface.

The Language Ecology Theory developed by Haugen (1972) offers a theoretic framework that is useful in explaining these processes since the language is regarded as a component of a system of forces, including social, cultural, and technological ones that interact. Though the authors by Mackenzie *et al.* (2020) use the concept of ecological thinking to determine the issue of digital inequality regarding AI design, they do not make explicit reference to African languages and voice technology. It demonstrates that there is a major gap in the literature:

no dedicated theory-led research that addresses the role of AI voice technologies in defining the sociolinguistic identity and digital survival of Hausa.

This paper fills this gap by applying sociolinguistic theory to the critical scholarly approach towards artificial intelligence to analyse how Hausa is positioned in the changing digital linguistic ecosystem.

Hausa Language in Sociolinguistic Context

The West and Central African language Hausa serves as the first tongue for more than 50 million individuals alongside being the second language for millions of people, based on Eberhard, Simons, & Fennig (2023) statistics. Hausa speakers predominantly reside throughout Nigeria and Niger, while additional communities use it in Chad and sections of Cameroon and Sudan. According to Bamgbose (2011), the expansion of Hausa was made possible through historical trade connections and the political leadership of the Sokoto Caliphate, leading to Sahel region growth.

Geographic Spread and Number of Speakers

Jibril (1982) describes Hausa language dispersion as extensive in northern Nigerian states and as the common language which people from various ethnic groups use to communicate. The language has a broad geographic presence in Niger, where it stands as the dominant tongue and extends further into rural and urban parts of Chad and Cameroon, according to the data presented by Eberhard *et al.* (2023). Hausa functions as the essential communication medium of the region, leading to its acceptance as a regional lingua franca.

Dialectal and Phonological Variations

Malmberg (1981) identifies Hausa dialects with separate phonological and lexical elements found throughout Kano, Katsina, and Sokoto, along with other regions in the language. The Kano dialect functions as the primary standard used in official activities yet different dialects from other parts of the country preserve their distinct linguistic patterns. In phonetic terms Hausa distinguishes word meanings by tone systems even though words share identical written forms. The way traditional tones are expressed differs extensively across the dialects, thus making it difficult for some speakers to understand each other (Jibril, 1982).

Hausa as a Marker of Identity (Ethnic, Regional, Religious)

Local identity in Hausa depends on all three aspects: ethnicity along with regional location and religious affiliation. Among African populations the Hausa stand as one of the largest ethnic groups which utilise their language as their main element for cultural identification. According to Bamgbose (2011), the unification of different ethnic sub-groups within the Hausa-speaking community takes place because of the powerful role of the Hausa language. Malmberg (1981) points out that the

dialects of different regions operate as solid markers of personal identity because native speakers identify strongly with their native dialect group above all else in the Hausa linguistic sphere.

The Hausa language maintains a deep religious connection with Islam because Islam introduced itself to the region during the 11th century. Through time many Arabic words entered the Hausa language, mainly for use in faith-related domains. The Hausa language functions as the main communication tool in Islamic education and religious sects in northern Nigeria and throughout Islamic education and religious practices, according to Oluwole (2013). The Hausa language functions as an essential tool through which Islamic teachings spread and operates as a religious identity marker for numerous users who speak the language.

Hausa in Education, Media, and Politics

The language of Hausa demonstrates importance which reaches beyond oral communication to education sectors as well as media sectors and political sectors. Hausa education is taught in primary facilities and used as an instructional language throughout numerous geographical regions of Nigeria, particularly in its northern regions. Bamgbose (2011) declares that this educational language prevalence ensures Hausa stands as a vital method for spreading knowledge and advancing intellectual growth.

The Hausa language stands at the forefront of Nigerian media outlets, which include all broadcast channels and several publications. Radio Kano, together with Arewa24, serves millions of Hausa-speaking communities throughout Nigeria and neighbouring territories. Oluwole (2013) explains that media platforms play a crucial part in promoting Hausa by establishing standardised usage patterns for modern communication which maintain its vitality in digital times.

Hausa stands as an essential political language throughout northern Nigeria because it serves as the main tool for both political speeches and campaign efforts and political discussion. Jibril (1982) points out political candidates depend on Hausa to communicate with numerous voters by delivering their political messages to broad constituencies efficiently. The Hausa language plays a fundamental role in politics, which makes it essential to both social and political settings in northern Nigeria.

Artificial Intelligence and Voice Technology: An Overview

Definitions and Functions of AI, ASR, and NLP

Artificial intelligence, through its systems, functions as a technology which develops digital capabilities to mimic human learning and problem-solving behaviours. Russell and Norvig (2016) describe how AI makes machines capable of executing operations which need human mental abilities. The essential part of automatic speech recognition (ASR) within artificial intelligence enables the processing of spoken language to create text through analysis of sound waves and association with phonetic

models. The article by Jurafsky and Martin (2020) points out how ASR technologies serve as an essential link that connects machine systems with human speakers during communications. Natural Language Processing (NLP) technology allows machines to translate human speech through ASR programs. NLP, according to Manning and Schütze (2003), provides machine users with sophisticated interactions through its text generation capabilities and sentiment analysis features.

Current Applications

According to Hoy (2021), AI solutions that incorporate Siri, which is a virtual assistant developed by Apple Inc., Google Assistant and Alexa demonstrate how ASR and NLP technology enhance regular communication between users and machines. Through the implementation of these technologies, Siri enables users to accomplish multiple tasks, including alarm setting and message sending. Google Assistant performs enhanced in-person response generation via its wider AI platform that processes user data. Through voice recognition integration with smart home devices, Alexa permits users to operate their environment by using their voice commands (Seitz, 2019). The increasing utilisation of AI across daily life activities can be observed through voice-operated systems that will keep improving due to advancements in ASR and NLP.

Global Linguistic Representation in AI Systems

The prevalence of voice technologies in society has increased yet global languages in AI programmes receive excessive attention to major global languages alone. According to Sager (2021), Google Assistant currently serves more than thirty languages but insists on overlooking several languages that have millions of speakers, including Hausa. AI design and development caters more strongly to dominant political and economic languages, as stated by Choudhury *et al.* (2020). Such limitations in representation create accessibility issues for speakers whose native languages, like Hausa, have minimal presence, whereas Hausa serves as an essential communication tool across African-speaking regions.

Existing Linguistic Hierarchies in AI Design

AI design incorporates linguistic dominance structures because the languages with more speakers and economic might control the development process of new technologies. The preference for widespread languages in technological development derives from the commercial business interests of companies that aim to address markets with higher economic potential (Mackenzie *et al.*, 2020). Bender *et al.* (2021) warn that economic models in technology development create system domination by large languages while important minority languages experience discrimination. The present design schemes of AI systems showcase social inequalities from society, which might intensify existing digital divisions by establishing linguistic and cultural alcoves.

Hausa and AI Voice Technology

Status of Hausa in Voice-Enabled Technologies

The local speech of West Africa exists as one of the most commonly used languages, with over 50 million native speakers, mainly in Hausa. Voice-enabled technologies only use Hausa in limited capacities. The dominance of English and Mandarin controls the AI voice system design while leaving behind languages like Hausa, according to Bender *et al.* (2021). Choudhury *et al.* (2020) report growing interest in AI linguistic expansion, which includes Hausa attempts for platform integration into Google Translate and other systems. These positive advancements occur at an early developmental phase since broader spoken languages lead the way.

Challenges in Integrating Hausa into AI Systems

Under-representation in Training Datasets

The integration of Hausa into AI systems faces an essential obstacle because training datasets are not extensive enough. Jurafsky and Martin (2020) emphasise how machine learning models need vast training datasets for effective operation in their explanations. Current AI systems face challenges processing the Hausa language because they lack sufficiently large and diverse datasets for this language. Mackenzie *et al.* (2020) identifies under-representation as a major problem since it hinders Hausa from achieving full integration into speech recognition or translation technologies.

Dialectal Complexity and Tonal Variation

Various parts of Hausa demonstrate ample dialectal diversity because pronunciation patterns and word choice and sentence structure rules differ among distinct regions. Automation of speech recognition faces major difficulties from language differences within its system, according to Russell and Norvig (2016). The mixture of dialectal variations and pronunciation tones makes spoken Hausa difficult for AI machines to process, which limits their ability to appear in voice-based technology solutions. The authors Jurafsky and Martin (2020) stress that automatic speech recognition systems primarily use standardised language training data so they miss essential dialectic language features.

Lack of Standardised Digital Corpora

The major barrier stems from the nonexistence of standardised digital libraries for the Hausa language. Any language needs standardised corpora consisting of precisely selected large text and speech datasets for effective AI system training. Standardised digital collections serve as essential requirements for AI systems to achieve successful processing and generation of language output. As Seitz (2019) indicates, the development of standardised digital corpora for all languages has been increasing, but the Hausa language suffers from this significant deficiency, which hinders

its adoption in voice technology systems.

Case Studies or Examples

Google Translate and Mozilla Common Voice

Google Translate provides limited translation capabilities for Hausa but its functionality remains restricted in relation to the French and Spanish languages. Poor translation accuracy between Hausa and other languages stems from inadequate translation data, which Hoy (2021) observes. The open-source voice dataset development of Mozilla's Common Voice project faces data shortages, especially for the language Hausa. Sager (2021) explains that although the Hausa language receives support from the initiative, its main obstacle is obtaining adequate linguistic data for training accurate AI models to perform ASR and NLP operations. Despite these challenges, both Google Translate and Mozilla Common Voice represent positive steps toward greater linguistic inclusivity in AI technologies.

Sociolinguistic Implications of Hausa's Under-representation

Identity Threats Pair with the Challenges of Obtaining Access as Well as Difficulties Providing Inclusion.

The absence of Hausa representation in artificial intelligence technologies creates important risks to maintaining the sociolinguistic identity of Hausa speakers. According to Bender *et al.* (2021), AI systems determine cultural identity through their language selection, especially for minority language users. Digital technologies that exclude Hausa speakers create marginalisation among their community by reducing the role of their language in modern times. The article by Choudhury *et al.* (2020) shows that digital platforms presently mould cultural engagement between users and their heritage, while ignoring Hausa disrupts the maintenance of both language and cultural identity. Lack of AI access in Hausa creates a vulnerability for their speakers to become marginalised from vital digital environments influencing social as well as economic and political opportunities.

Risks of Digital Marginalisation

Digital marginalisation risks becoming worse because Hausa is not present in AI technologies. According to Russell and Norvig (2016), digital systems replicate current social and linguistic tiers by endorsing the languages maintaining economic and political supremacy. Hausa faces disproportionate under-representation in world AI software that increases social inequality in its cultural territory. Seitz (2019) explains that massive AI platforms which disregard millions of speakers end up creating digital segregation that keeps these populations out of the global digital economy. This digital marginalisation restricts both information access and services while simultaneously suppressing the social and political strength of Hausa-speaking communities in modern digital times.

The System's Lack of Hausa Integration Results in Negative Consequences for Educational Functions and Cultural Preservation

Insufficient Hausa integration into AI systems creates disadvantages that extend into several areas, including communication systems as well as educational practices and cultural preservation. Jurafsky and Martin (2020) advocate for complete digital representation of languages since it maintains clear communication both educationally and professionally. The absence of Hausa in these systems prevents its use in schools and workplaces despite increasing technological adoption for learning and collaboration purposes. When AI systems omit the Hausa language, they also put cultural education heritage at risk of being lost. Mackenzie *et al.* (2020) demonstrate that language operates as a fundamental conduit for cultural heritage, whereas neglecting to support languages like Hausa in digital platforms results in cultural tradition disconnects which strongly impact younger people who focus on digital communication and education tools. The fading linguistic diversity in digital spaces works to discontinue cultural continuity within the Hausa community.

Strategies for Inclusion and Development Creating Hausa Voice Datasets

The initial essential move toward incorporating Hausa into artificial intelligence systems demands building extensive vocal database resources. Any language requires large, diverse datasets, according to Jurafsky and Martin (2020). Developing a proper Hausa voice dataset implies collecting various speech samples that capture the full spectrum of dialects and tones found in the language. Russell and Norvig (2016) explain that speech recognition models achieve their best performance through proper representation of data during training; therefore, Hausa needs sufficient presence in these datasets. The establishment of suitable datasets will improve the authenticity and dependability of voice-controlled technology for Hausa-speaking users while opening digital access for all.

Collaborations with Linguists, Developers, and Local Institutions

Local institutions, along with both developers and linguists, should work together effectively, as they play a crucial role in integrating Hausa into artificial intelligence systems. Language experts play a vital role when developing AI systems to guarantee accurate understanding of understudied languages, as Choudhury *et al.* (2020) explain regarding Hausa. Local universities and linguistic organisations should collaborate in order to create precise language models through their expertise and maintain Hausa AI systems that reflect its deep linguistic and cultural heritage. Bender *et al.* (2021) demonstrates how mutual alliances between Hausa experts and technology and AI development companies will guarantee voice technology emphasis on Hausa

while building applications which ensure comprehensive representation.

Open-Source Contributions (e.g., Crowdsourcing Transcriptions)

Open-source contributions prove to be a practical method to accelerate AI system development for Hausa. The Mozilla Common Voice platform provides Hausa native speakers with opportunities to donate their voices and transcriptions that contribute to training AI models through accessible datasets. According to Mackenzie *et al.* (2020), crowdsourcing enables fast data collection through community involvement, which creates a phenomenon of Hausa linguistic empowerment. Participation in such language representation projects enables communities to control their digital language presence and this will ultimately benefit their language conservation efforts and digital inclusion goals.

Policy and Advocacy for Linguistic Digital Rights

AI system representation of Hausa depends heavily on policy development and advocacy activities. The Washington University Graduate Journal of Legal Studies (2019) reminds us about linguistic digital rights, which need supporting policies to establish indigenous language usage in digital technologies. National and international policy frameworks designed for digital inclusion with priority for linguistic diversity should provide support for Hausa advocacy in digital platforms. The recognition of the Hausa language in AI systems by governments and policymakers needs to occur while they support programmes that protect native speaker language rights. Policies like these, according to Choudhury *et al.* (2020), help tech companies develop digital solutions for neglected languages.

RESULTS AND DISCUSSION

As the analysis shows, Hausa is still grossly under-represented among modern-day AI-powered voice technologies, even though it is geographically wide and has a high number of speakers. An overview of the voice-enabled platforms and AI language infrastructures reveals that Hausa is not represented at all, or only to a small degree, usually restricted to simple text-based programmes, but not to full-fledged speech recognition and speech synthesis. This is an imbalance of the larger global linguistic hierarchies in AI development, where the predominant global languages are prioritised in terms of commercial interests and availability of data.

The identification of the lack of data as one of the key limitations is one of the key consequences of the study. Existing Hausa speech data, which can be used to train Automatic Speech Recognition (ASR) systems, are also finite, disjointed, and not even spread across dialects. This has led to systems which do not reflect the phonological richness, tones and sociolinguistic diversity of Hausa. Ecologically, in the light of the Language Ecology Theory by Haugen, this kind of technological marginalisation

destroys the natural dynamics between language, society, and the changing digital space, which leaves Hausa at a disadvantage in the digital habitat.

The problems of orthographic and dialectal difficulties are also pointed out as obstacles to the introduction of AI. The standardised orthography of Hausa is not consistently used in electronic platforms, and the diversity of dialects is not taken into account in the design of AI. The effect of this is biased language modelling which favours some regional forms at the expense of others, which in turn cements internal linguistic hierarchies amongst Hausa speakers themselves. This has sociolinguistic implications for identity representation because the speakers might develop a lack of linguistic legitimacy in the online environment.

The other essential finding is the sociolinguistic implications of the digital exclusion. The lack of Hausa voice technologies prevents access to digital services, education and information for users who are non-literate or semi-literate and use spoken communication intensively. Such exclusion is potentially dangerous, as it contributes further to digital inequality and the loss of intergenerational linguistic transmission, especially among younger speakers, the linguistic habits of whom continue to be further influenced by technology.

To explain these outcomes, the research paper states that AI systems are not neutral technologies but rather sociotechnical mechanisms that mirror the relations of power inherent in the choice of language and priorities in the design. The under-representation of Hausa can be used to elucidate the situation when the technological development leads to a decline in language that occurs unconsciously in the ecological situation when the balance between both is not considered. The results, however, also show the availability of opportunities to become inclusive by creating databases collaboratively and developing open-source projects and advocating the use of linguistic digital rights with the help of policy.

All in all, the findings allow highlighting the necessity of the culturally responsive development of AI because Hausa is a living language system that requires complete representation in cyberspace. The Hausa voice technology integration is not only an engineering endeavour but also a sociolinguistic necessity that needs to be done to preserve diversity, identity, and fair access to linguistic resources in an era where artificial intelligence dominates.

CONCLUSION

This paper has discussed the overlapping of artificial intelligence, voice technology, and sociolinguistic identity with a particular focus on how the Hausa language can fit into the modern digital linguistic framework. By placing Hausa in the context of the global AI contexts, the study has shown that a high demographic presence does not necessarily result in a digital presence or an inclusion in the technological world. Although Hausa is one of the most actively used native languages in Africa, it is mostly not included in the large-scale AI voice technologies,

specifically speech recognition and voice assistant apps. The discussion determines structural causes that have led to this exclusion. Following the footprint of earlier research on speech and language processing, Jurafsky and Martin (2020) point out that the lack of training data, dialect differences, and orthographic inconsistencies limit the successful modelling of a language like Hausa. These technical pitfalls, however, transcend engineering constraints. Bender *et al.* (2021) warn that the exclusion of language in AI systems recreates wider sociocultural inequalities that endanger not only equal access to technology but also the symbolic significance of the excluded languages. The under-representation of Hausa in this respect is representative of an even greater trend of linguistic hierarchy in AI design practices.

The research also confirms that the linguistic representation of AI is not only a technical problem but also an ethical and sociopolitical matter. Choudhury *et al.* (2020) highlight that the presence of language in digital structures is a concern of digital rights and cultural sustainability. The result of excluding languages like Hausa as voice technologies means that speakers have limited access to information, education and online engagement, which further consolidates the already established marginalisation.

Notably, the research highlights the directions of the more inclusive online future. Raising awareness of collaborative models, Mackenzie *et al.* (2020) consider building large Hausa voice datasets with the help of community-based transcription campaigns and projects with linguists, technologists, and local institutions. Adding to this perspective, Russian and Norvig (2016) emphasise the importance of policy structures and advocacy in social work in streams of systemic bias in AI systems and linguistic diversity.

Overall, this study confirms that to achieve linguistic equity, cultural survival, and sustainable digital interaction and participation, Hausa must be integrated into AI voice technologies. This work also adds to sociolinguistic and AI research by preempting the indigenous language inclusion as an important aspect of ethical AI creation. Future studies can also look into the impact of tone, dialectal differences, and attitudes of speakers towards the Hausa speech recognition system and also look at comparative studies with other under-represented African languages. These questions will further inform the knowledge on the potential of AI to be utilised in service of, and not marginalising, linguistic diversity in the digital era.

REFERENCES

- Bamgbose, A. (2011). African languages today: The challenge of and prospects for empowerment under globalization. In E. Wolff (Ed.), *Languages and education in Africa: A comparative and transdisciplinary analysis* (pp. 1–17). Cambridge Scholars Publishing.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too

- big? *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–17. <https://doi.org/10.1145/3411764.3445773>
- Bird, S. (2020). Decolonising speech and language technology. *Proceedings of the 28th International Conference on Computational Linguistics*, 3504–3519.
- Birhane, A. (2020). *Algorithmic injustice: A relational ethics approach*. *Patterns*, 2(2), 1–12. <https://doi.org/10.1016/j.patter.2020.100205>
- Choudhury, M. D., Ang, D. S., & Park, K. (2020). Linguistic diversity in AI systems. *Journal of Linguistic Research*, 18(3), 35–49. <https://doi.org/10.1016/j.jlr.2020.06.001>
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2023). *Ethnologue: Languages of the world* (26th ed.). SIL International. <https://www.ethnologue.com>
- Haugen, E. (1972). *The ecology of language*. Stanford University Press.
- Jibril, M. (1982). *Hausa: An introductory analysis of the language*. University of Jos Press.
- Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed.). Pearson.
- Mackenzie, D., Cox, M., & Williams, B. (2020). Language inequality in artificial intelligence design. *Journal of Technology and Society*, 22(4), 87–95. <https://doi.org/10.1016/j.jts.2020.02.006>
- Malmberg, B. (1981). *Hausa: A sociolinguistic description*. Almqvist & Wiksell.
- Oluwale, J. (2013). The role of Hausa in Islamic education in northern Nigeria. *Journal of Islamic Studies*, 15(2), 134–145.
- Osei, Y. (2021). African languages and the AI revolution: A critical perspective. *Journal of African AI and Society*, 2(1), 28–39.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.