

Promoting Hausa Language via AI and Digital Tools

Sumayyat Ibrahim Gambo
Taraba State University, Jalingo

DOI: <https://doi.org/10.60787/jolan.vol6no1.467>

Abstract

This study investigates the representation, performance, and limitations of the Hausa language across selected Artificial Intelligence (AI) and digital platforms, with a view to identifying strategies for its more effective promotion and preservation. The study adopts a qualitative content analysis design, examining how ChatGPT and Google Translate process Hausa text, idiomatic expressions, and proverbs, alongside an analysis of AI-generated Hausa speech content circulating on social media platforms and a review of Hausa text-handling support in digital writing applications such as WordPad. The study is anchored on three theoretical frameworks, namely Language Preservation Theory, Language Education Theory, and Digital Humanities Theory, which together illuminate why accurate AI representation of Hausa matters for the language's continued vitality, how AI tools may support or undermine language learning, and what is culturally at stake when AI systems misrepresent Hausa linguistic and cultural content. The findings reveal that although Hausa has gained some visibility in digital environments, its integration into advanced AI systems remains limited and inconsistent: AI tools such as ChatGPT demonstrate clear potential for supporting Hausa language teaching and learning, yet they consistently struggle with complex grammatical structures, dialectal variation, and culturally embedded vocabulary, as illustrated by the recurring mistranslation of proverb terms. Similarly, AI-generated Hausa speech content on social media exhibits persistent tonal and pronunciation inaccuracies, reflecting the broader shortage of high-quality Hausa speech and lexical datasets available for AI training. The study recommends sustained collaboration among linguists, AI developers, government institutions, and Hausa-speaking communities to achieve meaningful integration of Hausa into AI technologies.

Keywords: *language technology, language preservation/revitalization, indigenous language documentation, natural language processing (NLP), digital inclusion*

1. Introduction

Language plays an essential role in defining identity, preserving culture, and transmitting knowledge across generations. In multilingual societies such as Nigeria, indigenous languages are vital tools of cultural continuity and national development.

However, the dominance of English and globalization-driven communication trends have significantly marginalized the use of local languages, including Hausa, within digital and educational domains. Scholars such as Aturamu (2024), Ukaegbu and Obiagboso (2025), and Haruna, Lame and Bappayo (2025) have emphasized that digital technology, when strategically applied, can strengthen indigenous language teaching, learning, and preservation.

The recent advancement of Artificial Intelligence (AI) and digital communication technologies has created new pathways for indigenous language revitalization. Tools such as ChatGPT, AI-driven translation systems, and digital writing platforms (e.g., WordPad) can be adapted to enhance Hausa language teaching, learning, content creation, and interaction. These technologies support language learners through automated tutoring, natural language generation, and speech recognition, fostering more accessible and engaging experiences.

Given this context, it becomes imperative to explore how AI tools can contribute to Hausa language promotion and preservation. This study, therefore, investigates how emerging AI technologies such as ChatGPT, alongside digital writing tools, can support Hausa language teaching, learning, cultural transmission, and digital inclusion. It draws from theoretical perspectives in language preservation, language education, and digital humanities to highlight how linguistic heritage can thrive within modern technological ecosystems.

1.1. Objectives of the Study

The main aim of this study is to examine the promotion and use of the Hausa language across emerging Artificial Intelligence (AI) and digital platforms such as ChatGPT and WordPad, and to explore their pedagogical and research potential for Hausa language teaching, learning and preservation.

To achieve this aim, the study is guided by the following specific objectives to:

- i. Identify whether the Hausa language is currently used or represented in selected AI and digital platforms.
- ii. Examine the challenges associated with using the Hausa language on AI-driven platforms.
- iii. Explore strategies for promoting the use of the Hausa language in AI platforms for pedagogical and research purposes.

2. Literature Review

2.1. Conceptual Review

2.1.1. Artificial Intelligence (AI) and Language Technology

Artificial Intelligence (AI) encompasses computational systems capable of simulating human cognitive tasks such as learning, language comprehension, generation, and reasoning (Akinsanya & Kareem, 2023). In the domain of language technology, AI supports applications like speech recognition, machine translation, and natural language processing (NLP), which are central to enabling machines to understand and produce human language. These technologies become especially important in multilingual societies where language access and digital inclusion are key to participation in the global information economy.

2.1.2. Indigenous Languages in the Digital and AI Era

Indigenous languages serve as embodiments of cultural identity and collective knowledge. However, in the digital era they often face marginalization due to technological dominance of major world-languages (Orji & Korie, 2024). AI platforms present both opportunities and risks: on one hand, they can support documentation, amplification and revitalization of indigenous languages; on the other hand, their effectiveness depends on the availability of data, appropriate resources, and culturally relevant design (Akinsanya & Kareem, 2023). Thus, AI-driven language technologies afford a new domain for indigenous language preservation, but require intentional adaptation to local cultural-linguistic realities.

2.1.3. The Hausa Language and Digital Inclusion

Hausa is one of Africa's most widely spoken indigenous languages, yet when it comes to advanced AI systems, its presence remains relatively limited (HausaNLP, 2023). Efforts such as the open-source platform "HausaNLP" highlight the need for culturally aware NLP tools that respect Hausa idioms, dialectal variation and socio-cultural context (HausaNLP, 2023). The integration of Hausa into AI systems therefore becomes a matter of digital equity: enabling Hausa speakers to engage with intelligent technologies in their own language enhances participation, cultural affirmation and digital inclusion.

2.1.4. AI and Indigenous Language Preservation

AI offers new mechanisms for language preservation via large-scale text and speech processing, machine translation, content generation and archiving (Oveh, Aziken & Atomatofa, 2025). When applied to indigenous languages, these technologies can accelerate documentation, extend accessibility and create interactive learning environments. Nevertheless, their success is contingent on building robust language corpora, addressing orthographic standardisation and ensuring the involvement of language communities (Akinsanya & Kareem, 2023; Inuwa-Dutse, 2025). In the Nigerian context, the recent launch of a multilingual large language model (LLM) supporting Hausa, Yoruba and Igbo languages, demonstrates a national commitment to this agenda (Bilali, 2025).

2.1.5. Digital Humanities and Cultural Representation

The field of Digital Humanities bridges technology with cultural scholarship, emphasising how computational tools can preserve, represent and analyse cultural expressions. In this framework, embedding indigenous languages like Hausa into AI systems aligns with a digital humanistic mission to protect linguistic diversity and cultural heritage. It shifts the focus from merely using technology to ensuring that technological progress reflects and respects local linguistic realities.

2.2. Theoretical Framework

This study is anchored on three interrelated theoretical foundations: Language Preservation Theory, Language Education Theory, and Digital Humanities Theory. These theories were selected because each addresses a distinct dimension of the relationship between Artificial Intelligence and the Hausa language that this study seeks to investigate: why accurate AI representation of Hausa matters for the language's continued vitality, how AI tools may support or undermine language learning, and what is culturally at stake when AI systems represent Hausa inaccurately. Read together, the three frameworks provide an interpretive lens through which the findings presented are subsequently discussed.

2.2.1. Language Preservation Theory

Language Preservation Theory holds that the survival of a language depends on its continuous and active use across multiple domains of communication, and that a language becomes endangered when such domains progressively narrow, particularly

when the language is excluded from domains of prestige and high visibility such as education, mass media, and digital technology (Fishman, 1991; UNESCO, 2019). Within the framework of this study, this theory accounts for the significance of the limited and inconsistent representation of Hausa on AI platforms reported in Section 4.1.1: where AI systems handle Hausa inadequately, the language risks losing ground in an increasingly important domain of everyday communication. The theory similarly accounts for the gravity of the lexical errors discussed in Section 4.1.5, since the substitution of inauthentic meaning for authentic cultural vocabulary in AI-generated output constitutes precisely the form of domain erosion the theory identifies as a precursor to language endangerment.

2.2.2. Language Education Theory

Language Education Theory, particularly as articulated in scholarship on technology-mediated and communicative language instruction (Warschauer & Healey, 1998), holds that language acquisition is most effectively supported through exposure to interactive, communicative, and contextually authentic input, and that digital tools may extend such input beyond the conventional classroom when appropriately designed. This framework accounts for the more favourable findings of the present study: ChatGPT's demonstrated capacity to explain grammatical structures, generate vocabulary exercises, and simulate conversational practice, as reported in Section 4.1.2, is consistent with the theory's claim that technology can meaningfully scaffold language learning. The same framework, however, equally anticipates the limitations identified in Sections 4.1.3 and 4.1.5, since communicative competence is contingent upon the accuracy of linguistic input; where AI-generated speech misrepresents tonal and consonantal patterns, or where AI-generated translation misrepresents idiomatic and proverbial meaning, the educational utility of the tool is correspondingly diminished.

2.2.3. Digital Humanities Theory

Digital Humanities Theory conceives of technology not merely as a neutral delivery mechanism for information but as a site at which cultural meaning is actively represented, interpreted, and potentially reshaped (Drucker, 2012; Berry, 2012). This framework bears most directly upon the proverb mistranslation findings reported in Section 4.1.5 and the speech-synthesis findings reported in Sections 4.1.3, 4.1.6, and 4.1.7. When an AI system renders a culturally specific term such as *kilisa* as "church,"

or flattens lexical tone in synthesized Hausa speech, the resulting error is not merely a technical shortcoming but a misrepresentation of a cultural artefact within a medium that is increasingly regarded as authoritative by its users. Digital Humanities Theory thereby supplies the normative basis for the recommendations advanced in Section 4.1.8, namely, that the responsible integration of Hausa into AI systems constitutes a humanistic obligation as much as a technical one, and that this obligation can be discharged only through sustained collaboration between technology developers and the communities whose language and culture are represented.

These three theories situate the findings of this study along a single explanatory continuum. Language Preservation Theory explains why the accuracy of AI's treatment of Hausa carries implications for the language's long-term vitality; Language Education Theory explains how AI tools may support or undermine language acquisition depending on the reliability of their output; and Digital Humanities Theory explains what is culturally at stake when AI systems misrepresent Hausa linguistic and cultural content. This integrated framework is applied throughout the interpretation of findings and the discussion.

2.3. Empirical Review

Empirical work relevant to this study can be arranged along a continuum, from broad investigations of digital technology and indigenous Nigerian languages in general, through AI-specific studies of other indigenous languages, to studies that engage the Hausa language directly, whether through general digital media or, more narrowly, through Artificial Intelligence platforms themselves. The review below follows this progression, moving from the most general studies to those most closely related to the present research.

Aturamu (2024) investigated the role of high-impact digital devices in facilitating the learning and sustainable development of indigenous Nigerian languages in general, motivated by the growing digital divide and the threat globalization poses to linguistic diversity. Using a descriptive survey design, the study sampled 150 participants, including language instructors, students, and educational technology experts, selected through stratified random sampling across various regions of Nigeria, and administered a self-structured questionnaire titled Digital Devices and Sustainable Development of Indigenous Nigerian Languages (DDSDINL), analyzed using descriptive statistics. The findings revealed that digital devices significantly enhance the efficiency and engagement of language acquisition, with respondents agreeing that

such devices and resources (e.g., apps, online courses) play significant roles in motivation, documentation, and interactive practice for indigenous languages, while also serving as effective tools for documenting and revitalizing endangered languages. The study further found that integrating digital solutions into language initiatives meaningfully enhances sustainability efforts by improving access to information, reducing reliance on physical resources, and supporting cultural heritage preservation. Based on these findings, the study concluded that digital devices hold clear potential for promoting linguistic diversity and cultural preservation, and recommended improved internet infrastructure and technological support within educational institutions, alongside government policies and funding to incorporate digital technologies into indigenous language teaching, learning, and preservation programmes. This study establishes the broad premise that digital technology can support indigenous language sustainability, but its focus on general educational devices, rather than intelligent systems, leaves open the question of how AI specifically might advance this agenda; the present study narrows that focus to Artificial Intelligence and, more specifically, to Hausa.

Eke and Adeyemi (2024) examined how digital media tools can be utilized to foster indigenous communication systems in Nigeria more broadly, drawing on cultural modernisation theory and media technological determinism theory to analyze the relationship between digital technology and traditional cultural practices. The study adopted a desk review design, systematically collecting, analyzing, and interpreting existing research, reports, publications, and other relevant documents on digital media and indigenous communication systems within the Nigerian context, with findings organized thematically according to the study's research objectives. The findings revealed that while traditional forms of communication remain vital within Nigerian communities, particularly in rural areas, their representation on digital media platforms remains limited and often misrepresented, a situation worsened by the digital divide and the dominance of Western-centric content online. The study also found that key barriers, including technological limitations, low digital literacy, and a lack of culturally sensitive platforms, hinder integration, while digital media simultaneously offers significant opportunities for visibility, community-driven content creation, and cross-community collaboration in cultural preservation. The researchers recommended that digital media platforms collaborate with cultural custodians, improve digital infrastructure and literacy in indigenous communities, and develop culturally sensitive platforms to ensure accurate representation in the digital space. While this study

extends the discussion from physical devices to digital media representation more broadly, it does not engage Artificial Intelligence directly; the next step in this review accordingly turns to studies that examine AI's treatment of indigenous languages specifically.

Orji and Korie (2024) explored the current status and prospects of the Igbo language within an AI-driven world and highlighting the risk of “digital language extinction” facing minority and indigenous languages amid the dominance of global languages such as English, French, and Mandarin in technology ecosystems. The study employed a thematic literature review approach, drawing on existing scholarship and reports to assess Igbo's penetration and representation across prominent AI applications, platforms, and NLP research efforts, while also reviewing digital preservation initiatives and the broader challenges facing the language. The findings revealed that Igbo remains largely excluded from prominent AI applications, with no support on major voice assistants such as Siri, Alexa, or Google Assistant, poor translation quality on Google Translate, and a scarcity of annotated Igbo datasets that significantly limits AI model performance for the language; contributing factors include minimal digital content, limited speech and text corpora, and inadequate keyboard support for Igbo's unique tones and characters. Based on these findings, the researchers concluded that strategic policy interventions are necessary to prevent “algorithmic exclusion” and recommended mandating standardized Igbo dataset contributions from major technology firms, fostering partnerships between universities and language boards, and enforcing platform localization regulations. This study demonstrates that the marginalization of indigenous languages within AI systems is not unique to Hausa, but rather reflects a structural pattern affecting Nigeria's indigenous languages more broadly; it also establishes a methodological precedent, direct assessment of AI platform behaviour, that the present study adopts in examining Hausa specifically, the language to which the review now turns.

Having established that AI-driven exclusion is a pattern affecting indigenous Nigerian languages generally, the review now turns to studies situated within Hausa itself, beginning with one that addresses digital communication without yet engaging AI directly. Shuaibu (2025) analyzed nontraditional writing practices in Hausa on Facebook, focusing on how creative language use, colloquial expressions, and digital vernaculars contribute to the destandardisation of Hausa orthography. Using a qualitative content analysis approach, the study randomly selected 30 posts, comments, and replies from a Hausa-language Facebook page and analyzed them for spelling,

syntactic, and morphological errors, as well as code-switching and loanword usage, drawing on Sociolinguistic Theory and Communication Accommodation Theory to explain why users depart from standard orthography in informal digital spaces. The findings revealed that Hausa social media users commonly rely on phonetic spellings, abbreviations, borrowed words, and informal expressions that prioritize speed and entertainment over grammatical accuracy, with the frequent omission of diacritics further complicating word recognition and comprehension for learners and educators. The study concluded that social media functions as both a threat to and a potential vehicle for preserving standard Hausa, recommending interventions such as awareness campaigns, digital Hausa keyboards, spell-checkers, and educational tools. This study is significant for the present research because the orthographic inconsistencies it documents, the omission of diacritics and Hausa-specific characters in informal digital writing, are precisely the kind of input irregularities that complicate AI processing of Hausa text, a connection the present study develops directly in its findings on orthographic variation. However, because Shuaibu's focus remains on human writing behaviour on social media rather than on AI systems themselves, the review turns next to studies that bring Hausa and AI together explicitly.

Ahmed, Cavus, Lawan, and Sani (2024) evaluated the effectiveness of a mobile application called “Hausar Kurma” in teaching English to Hausa-speaking hearing-impaired students in Nigerian special schools. Using a single-subject research design with a one-group pretest-posttest method, the study selected ten hearing-impaired primary school students from Zamfara State Special School, Gusau, who underwent eight weeks of training using the Android-based app, which employed AI-driven image classification to translate Hausa Sign Language into English. The findings showed a significant improvement in students' academic performance, with average scores rising from 35% on the pretest to 91% on the posttest, and a strong effect size confirming the intervention's impact, although usability testing revealed comparatively lower ratings for translation accuracy. The study concluded that Hausar Kurma is an effective and well-received tool for teaching English to hearing-impaired Hausa-speaking students and recommended its broader adoption alongside further development of sign-language translation tools. This study confirms that AI can be successfully deployed within a Hausa-speaking educational context, but its scope is necessarily narrow, a specialized assistive technology serving sign-language translation for a specific learner population, rather than an assessment of how mainstream AI platforms handle the Hausa language as a whole. The present study extends this line of inquiry by examining Hausa's

representation across mainstream, general-purpose AI platforms rather than a single specialized application.

Yakasai and Mukoshy (2025) assessed the pragmatic implications of Artificial Intelligence for language, learning, and knowledge construction, with a particular focus on Hausa, aiming to provide background knowledge on how AI can be leveraged to empower the Hausa language in education. Using a pragmatic framework, the paper examined the impact of AI on communication patterns, authorship, and agency in educational settings, while highlighting several implications of AI for Hausa, including AI's influence on communication styles, ethical issues and bias in Hausa AI models, the role of Google Translate in the standardization of Hausa, dialectal bias in Mozilla Common Voice, and the contribution of HausaGPT to cultural preservation. The paper concluded by calling for collaborative efforts among linguists, technologists, and policymakers to ensure that AI serves Hausa speakers effectively and responsibly. This study shares the present research's central concern, the intersection of Hausa and AI, but approaches it through a pragmatic, education-oriented lens addressing communication and authorship rather than through direct, systematic testing of AI platforms' linguistic accuracy. The present study accordingly occupies the remaining gap in this literature: it builds on Yakasai and Mukoshy's recognition that Hausa-AI bias and accuracy are live concerns, and on Orji and Korie's methodological precedent of direct platform assessment, to conduct a focused content analysis of how specific, mainstream AI tools such as ChatGPT process Hausa text, speech, and culturally embedded vocabulary, a gap not addressed by any of the studies reviewed above.

2.3.1. Research Design

This study adopts a qualitative content analysis design to investigate the representation, performance, and limitations of the Hausa language across selected Artificial Intelligence and digital platforms. A content analysis design was considered appropriate because the central concern of the study, namely, the accuracy and cultural appropriateness with which AI systems process Hausa text and speech, is most directly examined through systematic analysis of AI-generated output rather than through self-reported perception data. Data were accordingly derived from the researcher's structured interaction with selected AI tools and from native-speaker review of AI-generated Hausa content circulating on digital and social media platforms.

2.3.2. Sources of Data

Data for this study were drawn from three principal sources:

- i. Direct testing of Hausa-language expressions, sentences, and proverbs on ChatGPT and Google Translate;
- ii. Observation and linguistic analysis of AI-generated Hausa video content shared by selected content creators on Facebook, TikTok, Instagram, and YouTube Shorts;
- iii. Examination of Hausa text-handling capabilities within digital writing applications such as WordPad.

2.3.3. Procedure for Data Collection

The researcher conducted structured tests by inputting a range of Hausa expressions, sentences, and proverbs into ChatGPT and Google Translate, and systematically recording the resulting output. Proverb translations were cross-checked against native-speaker knowledge of Hausa idiom, figurative meaning, and cultural usage in order to identify points of inaccuracy. AI-generated Hausa videos produced by the identified content creators were reviewed for pronunciation accuracy, tonal realisation, rhythm, and overall linguistic naturalness, drawing on the researcher's native competence in Hausa phonology. WordPad was examined for the presence or absence of Hausa-specific orthographic and editorial support, including spell-checking, predictive text, and diacritic handling.

2.3.4. Method of Data Analysis

Data were analysed using qualitative content analysis, whereby AI-generated text and speech output was systematically compared against standard Hausa usage to identify recurring patterns of accuracy, inaccuracy, and cultural appropriateness. Findings were subsequently organised thematically in accordance with the study's three specific objectives, namely, the representation of Hausa on AI platforms, the challenges affecting its accurate processing, and strategies for promoting its more effective integration.

3. Findings of the Study

Limited Representation of Hausa Language in AI and Digital Platforms

This study found that Hausa has gained some visibility in digital environments; however, its integration into advanced Artificial Intelligence (AI) systems remains

limited compared with dominant global languages such as English, French, Chinese, and Arabic. Although some AI and digital tools recognize Hausa text, their ability to process the language accurately remains inconsistent. For instance, testing Hausa expressions on ChatGPT revealed that the platform can generate basic Hausa sentences, translate simple phrases, and explain Hausa vocabulary. However, difficulties occur when handling complex grammatical structures, idiomatic expressions, proverbs, cultural meanings, and dialectal differences. Expressions such as “Ruwa ba ya tsami” (literally “water does not become sour”) may receive literal rather than culturally appropriate interpretations because AI systems rely heavily on available training data. Similarly, Google Translate provides Hausa translation support, but translation quality varies depending on sentence complexity.

Table 1: Google Translation

Hausa Expression	AI Translation Ability
Ina kwana?	Correctly translated as “Good morning”
Na gode	Correctly translated as “Thank you”
Allah ya taimaka	Sometimes translated literally without conveying its cultural meaning

Digital writing applications such as WordPad also permit Hausa text entry but lack advanced Hausa language support such as:

- i. Hausa spell-checking;
- ii. grammar correction;
- iii. Hausa dictionaries;
- iv. predictive text;
- v. automatic correction of Hausa orthographic symbols.

Therefore, the study found that Hausa is present in digital spaces but has not yet achieved full technological integration.

AI Tools Have Potential for Hausa Language Teaching and Learning

The findings indicate that AI technologies provide considerable opportunities for improving Hausa language teaching and learning. AI-powered applications can assist learners through grammar explanations, vocabulary development, translation,

pronunciation practice, and interactive conversation. For example, ChatGPT can function as a Hausa learning assistant by:

- i. explaining Hausa grammar;
- ii. generating vocabulary exercises;
- iii. translating sentences;
- iv. assisting with writing practice;
- v. simulating conversations.

Example:

User: Translate “I am learning Hausa language.”

AI Response: Ina koyon harshen Hausa.

The study also observed that emerging educational platforms are beginning to incorporate AI-supported language learning, indicating that Hausa has the potential to become more digitally accessible.

Social Media AI Content Reveals Persistent Speech and Pronunciation Challenges

A major finding of this study is that although AI-generated Hausa content is increasingly available on social media platforms such as Facebook, TikTok, Instagram, and YouTube Shorts, the quality of Hausa speech generation remains below native-speaker expectations.

The researcher observed several AI-generated Hausa videos from content creators including:

- a) Abba Musa Ibrahim (AI-assisted classroom conversation),
- b) Abbank Nuhunks (AI-generated interview and dialogue),
- c) Young Alaji Animation (animated Hausa storytelling),
- d) Last B Mdy (AI-generated Hausa comedy animation).

These examples demonstrate that AI is increasingly being used to produce Hausa educational materials, animations, interviews, and entertainment. However, close observation revealed recurring linguistic weaknesses in the synthesized speech.



Figure 1: AI-generated Hausa classroom interaction shared on Facebook

demonstrating educational use of AI in Hausa communication



Figure 2: AI-generated Hausa podcast/interview showing synthetic Hausa speech

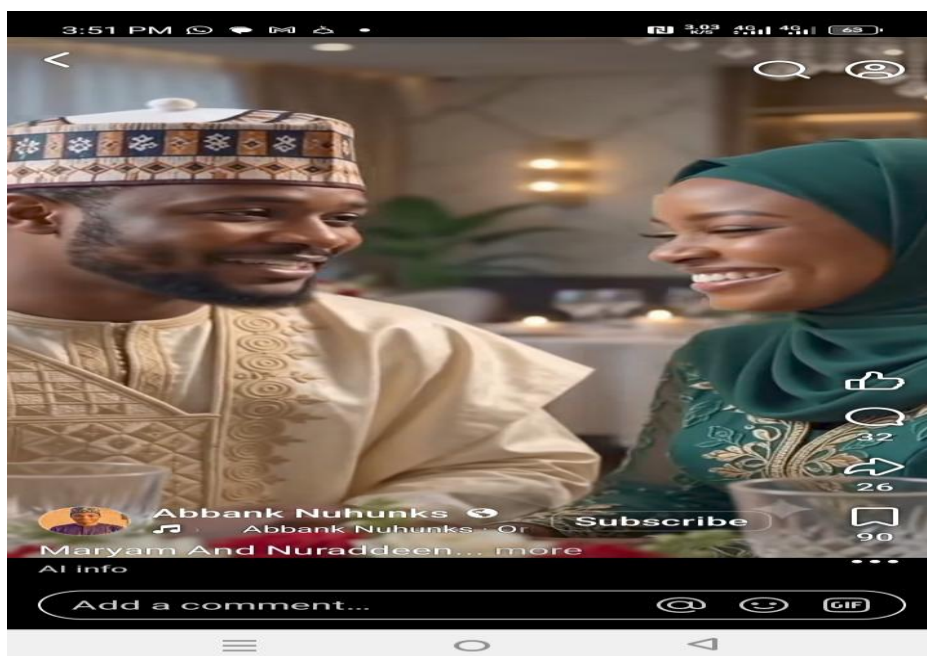


Figure 3: AI-generated Hausa dialogue between two virtual characters



Figure 4: AI-generated Hausa children's animation



Figure 5: AI-generated Hausa comedy animation distributed through Facebook Reels

The analysis of these AI-generated videos revealed several common problems:

- i. incorrect realization of Hausa lexical tone patterns;
- ii. inaccurate pronunciation of implosive consonants such as *b*, *d*, and *k*;
- iii. inconsistent pronunciation of ejective consonants;
- iv. unnatural rhythm and stress placement;
- v. English-influenced intonation patterns;
- vi. occasional mispronunciation of culturally familiar Hausa expressions.

Unlike AI speech systems developed for English, French, and Chinese, which generally produce natural and fluent speech because of the availability of extensive speech corpora and annotated pronunciation datasets, the Hausa versions often sound robotic and less natural. Native Hausa speakers can easily recognize these pronunciation inconsistencies, particularly in conversational contexts. This finding suggests that current AI speech synthesis systems have not yet been trained on sufficient high-quality Hausa speech data to accurately model native pronunciation and prosody.

Insufficient Hausa Language Data Limits AI Performance

The study discovered that the greatest limitation affecting Hausa AI performance is the shortage of high-quality digital language resources. AI systems require extensive datasets, including:

- i. written corpora;
- ii. recorded speech;
- iii. pronunciation databases;
- iv. dictionaries;
- v. annotated linguistic resources;
- vi. conversational datasets.

Compared with English, French, and Chinese, Hausa possesses significantly fewer publicly available digital resources for AI training. Consequently, current Hausa AI systems experience difficulties in:

- i. automatic speech recognition;
- ii. speech synthesis;
- iii. conversational AI;
- iv. grammar correction;
- v. sentiment analysis;
- vi. intelligent translation.

The pronunciation problems observed in the AI-generated Facebook videos further confirm the consequences of inadequate Hausa speech datasets. Most speech models have not yet learned native Hausa phonological patterns, resulting in unnatural voice output.

AI Mistranslation of Culturally Specific Hausa Proverb Vocabulary

A further illustration of the dataset limitations discussed in the fourth finding emerged when the researcher tested ChatGPT's ability to translate and explain a set of Hausa proverbs. While the platform produced broadly accurate translations and plausible thematic interpretations overall, closer examination by a native Hausa speaker revealed that the AI consistently misidentified or guessed incorrectly at certain culturally specific lexical items embedded within the proverbs, presenting these errors with the same apparent confidence as its correct output. Table 2 summarises three representative cases.

Table 2: AI Misidentification of Culturally Specific Vocabulary in Hausa Proverbs

Hausa Proverb	Misidentified Word	AI-Generated Meaning	Correct Meaning	Impact on Proverb Interpretation
A rashin lema ake kare ruwa da malafa .	Malafa	Calabash	Wide-brimmed straw hat used for protection from sun and rain	The proverb's logic depends on the substitute object plausibly replacing an umbrella. A wide-brimmed hat performs this function convincingly; a calabash (a bowl-shaped vessel) does not, weakening the imagery and practical sense of the saying.
Ban da shishigi, me ya gama talaka da kilisa ?	Kilisa	Church	Horse-riding undertaken for leisure/recreation, historically associated with the wealthy or noble class	The AI's reading reduces the proverb to a generic caution against meddling. The correct term reframes it as a class-conscious statement: a poor person has no legitimate business engaging in

				pursuits reserved for the affluent. The error likely arose from phonetic similarity to “kilisiya” (church).
Da tabaryar da ake daka, da ita ake lugude .	Lugude	Grinding (generic)	Rhythmic, coordinated pounding performed by several women sharing one mortar simultaneously	The generic reading reduces the proverb to a statement about a tool’s multi-purpose use. The correct term reveals a richer theme: the same resource accommodates both individual effort and cooperative, communal labour.

These findings reinforce the conclusion drawn in the fourth findings stated above: AI mistranslation of culturally embedded vocabulary is not a failure of general reasoning but a direct consequence of insufficient, fine-grained Hausa lexical data in the model’s training corpus. Where a term is rare, regionally specific, or absent from accessible digital dictionaries, the AI does not flag its own uncertainty; instead, it produces a fluent but inaccurate substitution, often by pattern-matching to a phonetically or contextually similar but semantically unrelated word, as seen with **kilisa/kilisiya** above. This poses a particular risk for proverbs and idiomatic expressions, where a single mistranslated term can distort the entire figurative meaning, the intended register (e.g., class commentary versus general social advice), and the cultural context in which the proverb is appropriately used. It further demonstrates that without native-speaker

verification, AI-generated translations of Hausa proverbs cannot currently be relied upon for academic, educational, or cultural documentation purposes.

Challenges of Orthographic Variation and Dialect Diversity

The study found that Hausa orthographic variation also creates challenges for AI processing.

Table 3: Hausa Orthography

Standard Hausa	Common Digital Form
Kauna	Kauna
ina son ka	ina sonka
yaya kake?	yaya kake

Many users omit Hausa-specific letters such as:

- i. k̄
- ii. d̄
- iii. ‘y
- iv. ɓ

Although these substitutions appear minor, they remove important phonological distinctions needed for accurate AI processing. Furthermore, dialectal variation among Kano, Sokoto, Katsina, Zazzau, and other regional varieties creates additional challenges for AI systems trained primarily on Standard Hausa.

Underrepresentation of Hausa in Speech Recognition Technologies

The findings further revealed that Hausa receives relatively little attention in modern speech recognition and voice assistant technologies. For example:

- i. English voice assistants understand natural speech with high accuracy.
- ii. French and Chinese speech recognition systems similarly perform well.
- iii. Hausa voice interaction remains limited, with lower recognition accuracy and less natural speech generation.

The AI-generated Facebook videos examined in this study demonstrate this limitation. Although the visual animations appear realistic, the accompanying Hausa speech often lacks native pronunciation, accurate tone, and appropriate conversational rhythm.

Improving Hausa speech technology would require:

- i. larger speech corpora;
- ii. recordings from native speakers across dialects;
- iii. phonetic annotation;
- iv. improved pronunciation models;
- v. collaboration between linguists and AI developers.

Digital Tools Encourage Hausa Content Creation and Cultural Preservation

Despite current limitations, the study found that digital technologies have significantly increased Hausa language visibility. The reviewed Facebook AI-generated videos demonstrate that content creators are using AI to produce:

- i. Hausa educational lessons;
- ii. animated storytelling;
- iii. interviews;
- iv. comedy skits;
- v. motivational content;
- vi. cultural narratives.

These digital productions contribute to:

- i. preserving Hausa language;
- ii. increasing language exposure among younger users;
- iii. promoting Hausa culture internationally;
- iv. encouraging digital literacy in indigenous languages.

However, linguistic quality assurance remains essential to prevent the spread of incorrect pronunciation and non-native speech patterns.

Need for Collaboration among Linguists, Developers, and Government Institutions

The study found that sustainable Hausa integration into AI requires collaboration among multiple stakeholders.

Table 4: Collaboration among Stakeholders

Stakeholder	Required Contribution
Linguists	Develop Hausa dictionaries, pronunciation guides, speech corpora, and grammatical resources
AI Developers	Build Hausa-compatible NLP, speech recognition, and text-to-speech systems
Government	Fund indigenous language technology and digital language policies
Universities	Conduct AI and Hausa language research
Hausa-speaking Communities	Provide authentic speech recordings, cultural expressions, and dialectal data

The findings demonstrate that technology alone cannot preserve Hausa. Meaningful AI integration requires sustained institutional support, high-quality linguistic resources, and active participation from native Hausa speakers.

5. Conclusion

The study concludes that Hausa has entered the AI and digital ecosystem but remains a low-resource language. The analysis of AI-generated content on Facebook and other social media platforms demonstrates that AI can successfully generate Hausa educational materials, animations, interviews, and cultural content. Nevertheless, the synthesized speech frequently exhibits inaccurate tone patterns, incorrect pronunciation of Hausa phonemes, and unnatural prosody compared with AI systems for English, French, and Chinese. The mistranslation of culturally specific proverb vocabulary, discussed in Section 4.1.5, further demonstrates that current AI systems lack the fine-grained lexical and cultural data needed to process Hausa idiomatic expression reliably. These findings underscore the need for larger Hausa speech datasets, culturally informed AI models, and stronger collaboration among linguists, AI developers, government institutions, and Hausa-speaking communities to ensure that AI technologies support the accurate preservation, teaching, and global promotion of the Hausa language.

References

- Ahmed, L., Cavus, N., Lawan, A. A., & Sani, I. (2024). Hausar Kurma: Development and evaluation of interactive mobile app for the English–Hausa sign language alphabet. *IEEE Access*, 12, 46011–46025.
- Akinsanya, A., & Kareem, F. T. (2023). Artificial intelligence (AI) in multilingual education: Investigating its potential in supporting the preservation and instruction of indigenous languages in formal education. Kwasu. <https://kwasuspace.kwasu.edu.ng/handle/123456789/3347>
- Aturamu, O. O. (2024). Leveraging high-impact digital devices to facilitate the learning and sustainable development of indigenous Nigerian languages. *Journal of Liaoning Technical University (Natural Science Edition)*, 18(8), 328–340.
- Berry, D. M. (2012). Introduction: Understanding the digital humanities. In D. M. Berry (Ed.), *Understanding digital humanities* (pp. 1–20). Palgrave Macmillan.
- Bilali, H. (2025, September 22). Nigeria launches N-ATLAS, an AI language model in Yoruba, Hausa, Igbo and Nigerian English. *We Are Tech*. <https://www.wearetech.africa/en/fils-uk/news/tech/nigeria-launches-n-atlas-an-ai-language-model-in-yoruba-igbo-hausa-and-nigerian-english>
- Drucker, J. (2012). Humanistic theory and digital scholarship. In M. K. Gold (Ed.), *Debates in the digital humanities* (pp. 85–95). University of Minnesota Press.
- Eke, C., & Adeyemi, M. O. (2024). Utilising digital media tools to foster indigenous communication systems in Nigeria. *African Journal of Social and Behavioural Sciences (AJSBS)*, 14(6), 3303–3322.
- Fishman, J. A. (1991). Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages. *Multilingual Matters*.
- Haruna, A. S., Lame, A. S., & Bappayo, B. (2025). The role of artificial intelligence in promoting Hausa culture, literature and translation. *International Journal of African Development and Sustainable Research*, 7(2). ISSN: 3026-8222.
- HausaNLP. (2023). About HausaNLP. Bayero University, Kano. <https://hausanlp.org/about.html>

- Inuwa-Dutse, I. (2025). NaijaNLP: A survey of Nigerian low-resource languages [Preprint]. arXiv.
- Orji, D.-M. A., & Korie, I. J. (2024). The Igbo language in an AI world: Assessing its current status. *OHAZURUME – Unizik Journal of Culture and Civilization*, 3(1), 137–145.
- Oveh, R. O., Aziken, G., & Atomatofa, E. (2025). Leveraging AI and digital technologies for cultural preservation and exchange in Nigeria. *Journal of Science Research and Reviews*, 2(1). <https://doi.org/10.70882/josrar.2025.v2i1.48>
- Shuaibu, A. (2025). Hausa language on Facebook: A linguistic analysis of grammatical errors and their effects on orthography. *Journal of Second and Multiple Language Acquisition (JSMULA)*, 13(1), 730–749. <https://doi.org/10.5281/zenodo.15570333>
- Ukaegbu, N. M., & Obiagboso, C. (2025). Exploring the use of digital resources and tools in improving Igbo language learning in Nigerian universities. *Nsukka Journal of the Humanities*, 33(1). <http://www.njh.com.ng/>
- UNESCO. (2019). *Indigenous languages in the digital age: Opportunities and challenges*.
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57–71.
- Yakasai, S. A., & Mukoshy, J. I. (2025). Exploring pragmatic implications of artificial intelligence on language, learning, and knowledge construction: The case of Hausa. *JOLAN Supplement*, Volume V(1), 204–215. <https://doi.org/10.60787/jolan.vol5no1.419>

Sumayyat Ibrahim Gambo
Department of Languages and Linguistics
Taraba State University, Jalingo
+2348031322451/+2349045724091
Pga06326sumayyat@gmail.com & isumayyat@yahoo.com