

Exploring pragmatic implications of artificial intelligence on language, learning, and knowledge construction: The case of Hausa

DOI: <https://doi.org/10.60787/jolan.vol5no1.419>

Salisu Ahmad Yakasai

Department of Nigerian Languages, Usmanu Danfodiyo University, Sokoto
syakasai2013@gmail.com

Jamilu Ibrahim Mukoshy

Department of Nigerian Languages, Usmanu Danfodiyo University, Sokoto
jmukoshy@gmail.com

Abstract

This paper examines the pragmatic implications of Artificial Intelligence (AI) for language, learning, and knowledge construction, with a particular focus on Hausa. Its aim is to provide background knowledge on how AI can be leveraged to empower the Hausa language and education. Using a pragmatic framework, the paper analyzes the impact of AI on communication patterns, authorship, and agency in educational settings, while also exploring the potential benefits and challenges of AI integration in education and research. Specifically, the paper highlights several implications of AI for Hausa, including: AI's influence on communication styles and expressions; ethical issues and bias in Hausa AI models; AI's impact on cultural identity and linguistic evolution; the role of Google Translate in the standardization of Hausa; dialectal bias in Mozilla Common Voice; the contribution of HausaGPT to cultural preservation; and YouTube's moderation of Hausa content. It also discusses the evolving role of AI in Hausa knowledge construction and cultural heritage preservation. Finally, the paper offers recommendations and suggests areas for further research. It calls for collaborative efforts among linguists, technologists, and policymakers to ensure that AI serves Hausa speakers effectively and responsibly.

Keywords: Hausa language, pragmatics, artificial intelligence, language learning, knowledge construction

Introduction

AI is increasingly playing a transformative role in language, learning, and knowledge construction, with specific applications emerging for languages like Hausa. In the context of Hausa, AI is being utilized to bridge educational gaps and preserve cultural heritage. AI-powered tools are facilitating personalized language learning experiences, enhancing engagement through interactive platforms, and improving accessibility for Hausa speakers. Notably, initiatives like the collaboration between Google, the Kaduna State Government, and Data Science Nigeria are producing AI education materials in Hausa, demonstrating a commitment to inclusive technological advancement. These efforts aim to democratize AI knowledge and empower Hausa speaking communities.

Furthermore, AI contributes to the preservation and dissemination of Hausa cultural and literary works. Machine translation and natural language processing (NLP) technologies enable the digitization and wider distribution of Hausa literature, ensuring its accessibility to global audiences. AI's capacity to analyze linguistic patterns and trends also supports the study and preservation of the Hausa language itself. As highlighted by research, AI presents a powerful opportunity for preserving and promoting Hausa culture, language, and literature through machine translation, speech recognition, and AI-generated content (Mbuk, 2021). The development of "Hausa AI" through NLP, machine learning, and speech recognition technologies allows for enhanced communication and interaction for Hausa speakers, with applications in education, customer support, and content creation (Lingvanex). These advancements signify the growing importance of AI in supporting and strengthening the Hausa language and its cultural significance.

Aim and objectives of the paper.

This research aims to provide a background on how to leverage AI to empower Hausa language and educational sustainability.

The key objectives are as follows:

- i. Investigate how AI tools (e.g., NLP models, speech recognition, and machine translation) can enhance the documentation, standardization, and revitalization of Hausa.
- ii. Examine challenges such as limited datasets and dialectal variations in developing effective AI solutions.

- iii. Analyze the effectiveness of AI-powered platforms (e.g., adaptive tutors, chatbots, and TTS systems) in improving literacy and STEM education for Hausa speakers.
- iv. Identify gaps in culturally relevant AI educational content and propose localization strategies.

Literature review

The Hausa language, spoken by over 80 million people across West Africa, holds a prominent position in African linguistics but remains underrepresented in digital technologies and AI applications. As Artificial Intelligence (AI) and Natural Language Processing (NLP) expand globally, there is growing interest in leveraging these technologies for language preservation, educational development, and knowledge democratization among Hausa speaking communities.

Efforts to preserve and digitize Hausa have accelerated through initiatives such as the HausaNLP project, an open source community committed to developing language resources and fostering collaboration for NLP applications in Hausa. Recognizing Hausa as the second most spoken language in Africa, this initiative addresses the scarcity of computational tools and data. Similarly, AfriDataHub has curated NLP datasets specifically for Hausa, supporting the creation of parallel corpora that enhance AI model training and ensure cultural relevance in linguistic technologies.

Additional developments include the creation of AI-powered text-to-speech (TTS) models, such as those by the AI Bauchi Community, which seek to make Hausa more accessible through natural-sounding, machine-generated voice outputs. These technologies not only promote language preservation but also expand the usability of Hausa in digital environments. However, challenges remain; as Haruna et al. (2020) observe, many models lack dialectal diversity and are based on formal registers, limiting their adaptability across Hausa speaking populations.

The educational potential of AI is particularly relevant in Northern Nigeria, where language barriers often restrict access to quality instruction. In collaboration with the Kaduna State Government and Data Science Nigeria, Google launched the AI for Beginners Learning Video Series, an initiative that presents foundational AI concepts in Hausa to broaden participation in tech education (Independent, 2024). This project exemplifies the value of culturally and linguistically tailored content for learner engagement and comprehension.

Complementing this, Garba and Bello (2022) emphasize the need for AI driven tutoring systems in underserved regions, particularly where teacher shortages hinder literacy and STEM education. Ismaila et al. (2023) add that Hausa language chatbots could serve as interactive learning agents, yet their effectiveness hinges on culturally adaptive dialogue and locally relevant content. The Hausa Intelligence Chatbot System (Abdullahi & Sadiq, 2021) represents a promising step toward conversational AI in Hausa, offering user friendly platforms that foster inclusivity in learning and information access.

AI offers a powerful platform for democratizing knowledge, but its success in Hausa speaking contexts depends on ethical and inclusive implementation. Tools such as Google Assistant and Siri currently do not support Hausa, thereby excluding non-literate and rural users who rely heavily on spoken interaction. The African AI Observatory (2023) notes that many AI systems, when developed without community involvement, risk reinforcing linguistic hierarchies and digital inequity.

Community led projects, such as the Hausa Wikipedia Initiative and HausaNLP, illustrate how open source and participatory models can empower local knowledge creation. However, without institutional backing and sustainable infrastructure, these initiatives face long term viability issues. As Bird (2020) asserts, preserving indigenous languages through AI requires not only technological tools but also policy frameworks that protect linguistic diversity and promote digital equity.

Despite promising developments, several gaps remain in Hausa AI research and application. First, dialectal inclusivity is critical; existing models tend to prioritize standard Hausa, sidelining regional variants. Second, educational technologies must move beyond translation to incorporate culturally responsive pedagogy. Lastly, ethical concerns around data ownership, representation, and bias necessitate the development of decentralized, community driven AI policies.

Pragmatic implications of AI on Hausa language, learning, and knowledge construction

Pragmatics, the study of language in context, provides a useful framework for analyzing AI's impact on communication. Artificial Intelligence (AI) is transforming the educational landscape, offering innovative solutions for teaching, learning, and research. However, the

integration of AI raises important questions about language, communication, and knowledge construction. Here are some key concepts:

AI's influence on Hausa communication styles and expressions

The integration of Artificial Intelligence (AI) into Hausa language platforms is reshaping communication norms, particularly in digital and social media spaces. AI driven tools such as machine translation (MT) systems - Google Translate, for example are central to these shifts. These tools, though groundbreaking, frequently misrepresent Hausa idiomatic expressions. For instance, Google Translate often translates English idioms like "break a leg" into overly literal Hausa forms ("karya kafa"), omitting the cultural connotation encapsulated in the phrase's equivalent "Allah ya ba da sa'a" (Haruna et al., 2020). This highlights how AI's syntactic and semantic limitations fail to capture the depth of meaning embedded in the Hausa language.

Moreover, AI-generated content such as chatbots, social media posts, and even automated news updates tends to favor standardized, formal Hausa. This leaves colloquial dialects like Dauranci or Katsinanci marginalized in everyday communication. Research by Adelani et al. (2021) indicates that AI systems trained on formal texts, such as news articles, struggle with the more dynamic, informal registers used in daily Hausa discourse, leading to a lack of nuance and cultural context. As younger generations increasingly rely on these AI mediated interactions, they risk adopting these standardized language forms, which could erode the richness of oral traditions, such as proverbs (karin magana) and context dependent humor, which are central to Hausa communication. Mbuk (2021) stresses the need for AI developers to integrate colloquial variants to preserve the authenticity of these linguistic practices.

Ethical issues: Bias in Hausa AI models

The biases inherent in AI systems especially those trained on limited or non-representative datasets pose significant ethical challenges for Hausa speaking communities. One notable example is the Mozilla Common Voice project. Initially, its speech recognition system had an inadequate representation of rural dialects and female voices, leading to a high error rate among rural Hausa speakers (Abdullahi et al., 2022). This problem exemplifies how AI models, if not carefully curated, can perpetuate digital inequalities by failing to understand diverse regional accents and pronunciations. Similarly, AI content moderation systems, such as those used by platforms like YouTube, have struggled to contextualize and

moderate Hausa language content accurately. A study by the African AI Observatory (2023) highlighted that such systems disproportionately flagged Hausa language videos, including educational content, as "spam" or "violent," due to cultural misunderstandings or lack of nuanced data. This raises concerns about digital exclusion, as Hausa speakers may avoid using these platforms due to misclassification of their content or the inability to access appropriate information.

In response, researchers like Ismaila et al. (2023) have called for participatory dataset curation, where native speakers and cultural experts audit and correct these AI biases. This approach can help reduce errors in speech recognition and content moderation, thus improving AI accessibility for all Hausa speakers, regardless of gender, dialect, or region.

AI's impact on cultural identity and linguistic evolution

While AI holds promise for the preservation of Hausa heritage through initiatives like digital archives and the documentation of oral traditions, it also accelerates the linguistic shifts within the community. A key manifestation of this change is code switching. AI models, particularly those used for translation, frequently introduce code switching between Hausa and dominant colonial languages such as English and French, reinforcing the linguistic hierarchies that have historically marginalized indigenous languages (Garba & Bello, 2022). For instance, AI driven translation systems often render terms like “internet” or “download” into their English forms rather than using locally established terms like "intanet" or "sauke".

Projects like HausaGPT, an open source language model, aim to preserve Hausa linguistic identity by generating culturally and contextually relevant content, such as proverbs (karin magana) and poetry (wakoki). In tests, 78% of native speakers rated its generated outputs as culturally authentic (Garba et al., 2024). However, even such projects have encountered challenges, HausaGPT occasionally uses English loanwords in place of native terms, reflecting broader trends in linguistic shift. The Kano AI Lab, which focuses on dialect preservation, exemplifies a community driven approach to balancing technological progress with the preservation of linguistic diversity (Lingvanex, 2024). This highlights the ongoing tension between standardizing Hausa for global usability and preserving its rich, multifaceted dialects.

Google translate and the standardization of Hausa

The introduction of Hausa into Google Translate in 2019 marked a significant leap toward digital accessibility for Hausa speakers. However, studies have shown that while the tool aids in communication, its output often misrepresents colloquial expressions. A 2023 study found that only 62% of informal Hausa phrases were translated correctly by Google Translate, compared to 89% for formal texts (Adelani et al., 2023). This highlights how AI models, while effective for standard language, may contribute to the marginalization of dialects and expressions rooted in everyday Hausa communication.

Mozilla common voice and dialectal bias

Mozilla common voice, a crowd sourced platform for training speech recognition models, initially faced significant challenges in representing the full spectrum of Hausa dialects. Its 2021 dataset contained 85% of recordings from urban Kano, leaving out rural dialects like Sokoto Hausa (Abdullahi et al., 2022). This resulted in higher error rates - up to 40% when recognizing speech from rural or non-native speakers. However, after advocacy by linguistic groups such as Masakhane, Mozilla collaborated with Nigerian universities to gather a more balanced dataset that included diverse dialects. By 2023, this improvement had boosted recognition accuracy by 25% (Adelani et al., 2023).

HausaGPT and cultural preservation

The development of HausaGPT represents an exciting application of AI to strengthen cultural identity. This open source project, trained on classical texts like *Ruwan Bagaja* and *Magana Jari Ce*, can generate proverbs, riddles, and praise poetry in Hausa. Notably, 78% of native speakers tested rated the model's output as culturally authentic (Garba et al., 2024). However, despite its success in preserving cultural content, challenges remain: HausaGPT sometimes incorporates English loanwords (e.g., "downloading" instead of "saukewa"), illustrating the ongoing tension between AI's global reach and the evolving nature of indigenous languages (Lingvanex, 2024).

YouTube's moderation of Hausa content

YouTube's AI-powered content moderation system has often misclassified Hausa language videos. In a 2022 study, researchers found that 30% of videos related to Boko Haram were wrongly flagged as "terrorist content" due to the system's failure to interpret the sarcastic, satirical, and humorous tones prevalent in Hausa discourse (zambo). Similarly, comedic skits using satire were flagged as "hate speech" (Ismaila et al., 2023). To bypass these filters, content creators have resorted to adding English subtitles, thereby diluting the

linguistic authenticity of their videos. This highlights the need for more contextually aware AI systems that understand the nuances of Hausa communication.

The evolving role of artificial intelligence in Hausa knowledge construction and cultural preservation

Artificial Intelligence (AI) is playing a transformative role in the construction, organization, and dissemination of Hausa knowledge systems. Through technologies such as Natural Language Processing (NLP), automated summarization, and information retrieval, AI enables more efficient interaction with Hausa language resources. For instance, text summarization algorithms like TextRank and BERT based models have been applied to condense voluminous historical works such as *Tarihin Kano* and *Magana Jari Ce* into accessible summaries for educational purposes (Adelani et al., 2021). This capability helps students and researchers engage with complex texts more efficiently, especially within digital learning environments. Similarly, AI driven information retrieval systems, such as Elasticsearch engines trained on Hausa corpora, allow users to locate specific themes or keywords in digitized collections of (littattafan Hausa) Hausa books, thereby promoting structured digital archiving of literature, historiography, and traditional knowledge (Garba & Bello, 2022).

A particularly vital application of AI is in the preservation and analysis of Hausa oral traditions such as proverbs (karin magana), folktales (labaran gargajiya), and indigenous songs (wakoki). Speech recognition tools, including those developed by the Masakhane and HausaNLP collectives, have begun transcribing spoken Hausa across various dialects, facilitating the preservation of oral genres traditionally passed down through generations (Masakhane, 2023). For example, AI tools have been used to digitize proverbs such as "Ruwa ba ya tsami a banza" ("Water does not become sour for nothing"), enabling their categorization by theme, usage context, and rhetorical form. NLP models are further employed to extract patterns from these texts, mapping metaphorical structures, similes, and idiomatic expressions used across regions such as Kano, Katsina, and Zaria (Yahaya, 2020).

In addition, AI models like HausaGPT, trained on Hausa literature and oral archives, have demonstrated the ability to generate culturally relevant content such as praise poetry (wakokin yabo), children's stories (labaran yara), and interactive riddles (kacici-kacici), with 78% of native speakers rating the outputs as culturally authentic (Garba et al., 2024).

Projects such as AfriNLP and the Kano AI Lab dialect preservation initiative have further emphasized inclusive data curation by integrating underrepresented dialects like Sakkwatanci and Dauranci (Lingvanex, 2024).

Nevertheless, the generative power of AI presents ethical challenges. Many machine translation tools, including early versions of Google Translate and Lingvanex, have misrepresented complex Hausa expressions due to lack of contextual grounding. Moreover, there are risks of historical inaccuracies when AI systems generate narratives about cultural or religious events without adequate training on authentic sources. The Sauti chatbot, developed by students at Ahmadu Bello University, has shown success in delivering Hausa language advice on health and agriculture, but its limitations in dialect handling highlight the need for localized training and human in the loop validation (Ismaila et al., 2023).

To ensure cultural integrity and linguistic accuracy, there is an urgent need for collaborative, interdisciplinary frameworks. Linguists, software engineers, oral historians, and cultural custodians must co-develop culturally aware datasets and validation protocols. Initiatives like the AfriDataHub Hausa Repository exemplify how community informed AI development can foster authentic representation and sustainable knowledge ecosystems. Furthermore, the establishment of benchmark corpora for Hausa, similar to what exists for English and French would greatly improve the evaluation of AI outputs across domains such as education, entertainment, and policymaking (Bird, 2020).

In sum, AI holds significant promise in preserving and enriching Hausa linguistic and cultural heritage. From digitizing oral traditions to creating interactive learning tools, AI can bridge generational gaps and promote access to indigenous knowledge. However, realizing this potential demands ethical stewardship, inclusivity, and respect for linguistic diversity to prevent the dilution or distortion of cultural values.

Conclusion

Artificial Intelligence (AI) is significantly transforming Hausa language learning and knowledge construction by introducing practical, scalable solutions to long standing educational and cultural challenges. Pragmatically, AI enhances access to educational resources, especially for learners in remote or underserved areas, through the digitization and semantic indexing of Hausa texts, folktales, proverbs (karin magana), and oral histories. These digitized materials converted through tools like optical character recognition (OCR), automated summarization, and NLP - form searchable archives that

promote inclusive access to indigenous knowledge. For instance, automated transcription and summarization of classical works like *Tarihin Kano* or *Magana Jari Ce* allow educators and researchers to efficiently extract relevant information, facilitating integration into modern curricula.

Machine translation systems and cross lingual NLP models are also reducing linguistic barriers, enabling global audiences to engage with Hausa literature and cultural expressions. Tools like Google Translate, Lingvanex, and experimental Masakhane multilingual models have opened access to Hausa content in international research and media. Moreover, AI supports personalized language learning through adaptive platforms that tailor content to users' proficiency levels. Interactive AI-powered tutors and chatbots, such as those developed by Data Science Nigeria and AI Hub Nigeria, offer instant feedback on grammar, pronunciation, and usage. These tools help learners acquire the language more effectively while fostering engagement through gamified experiences. Collaborations with global tech platforms, including Google's AI for Africa initiative, have resulted in the creation of Hausa language educational materials, promoting both technological literacy and language preservation among Hausa speaking populations.

Recommendations

To fully harness AI's potential in Hausa contexts, it is imperative to prioritize the development of culturally grounded and linguistically accurate systems. This begins with fostering interdisciplinary collaborations between AI engineers, Hausa linguists, oral historians, and cultural custodians. Such partnerships can ensure that AI-generated content accurately reflects the structural and semantic nuances of the language. For example, validation mechanisms such as human in the loop feedback and culturally informed benchmarking can help detect errors or cultural mismatches in AI-generated Hausa content.

Developing localized and dialect sensitive datasets capturing dialectal varieties like Zazzaganci, Sakkwatanci, and Kananci is essential to improving accuracy and relevance. Current models often struggle with dialectal diversity due to limited training data. Community led data curation projects, like those seen in the AfriNLP and Masakhane communities, can serve as a blueprint for building inclusive Hausa corpora.

Furthermore, investment in open source Hausa NLP tools, including automatic speech recognition (ASR), sentiment analysis, and text to speech (TTS) systems, can democratize access and stimulate innovation. Platforms such as Hugging Face now host emerging Hausa

models, but sustained support is needed to refine and expand these tools. Parallel to technical development, public digital literacy initiatives should empower Hausa speaking communities to understand, use, and critically evaluate AI applications. Targeted workshops and educational campaigns can promote ethical AI use in language learning, content creation, entrepreneurship, and governance.

Areas for further research

Future research must address the linguistic and ethical dimensions of AI integration in Hausa contexts. Technically, more sophisticated NLP models capable of handling the agglutinative morphology, tonal shifts, and pragmatic markers characteristic of Hausa are needed. Research into prosody aware TTS and robust grammatical parsers could enhance the authenticity of machine generated Hausa speech and text. Moreover, the development of generative AI systems capable of composing culturally faithful narratives, poetry (waka), and proverbs should be guided by in depth ethnolinguistic studies.

Preserving endangered dialects and oral traditions through AI is another critical research frontier. Many undocumented variants of Hausa, spoken in rural or border communities, are at risk of disappearing. AI tools such as dialect recognition systems, neural transcription, and AI driven storytelling platforms could serve as digital repositories for these linguistic and cultural treasures.

Equally important are investigations into the ethical implications of deploying AI in Hausa societies. Concerns around algorithmic bias, data privacy, misinformation, and cultural misrepresentation require urgent scholarly attention. Frameworks for responsible AI rooted in local values and inclusive governance should guide AI use in education, media, and governance. Finally, exploring the integration of AI in Hausa educational systems through intelligent tutoring systems, adaptive assessments, and curriculum development can significantly improve learning outcomes and support lifelong learning within Hausa speaking communities.

References

Abdullahi, M., & Sadiq, M. (2021). Hausa intelligence chatbot system. In A. S. Al-Fuqaha et al. (Eds.), *Advances in Science, Technology & Innovation*. (pp. 195-204). Springer. https://doi.org/10.1007/978-3-030-90618-4_19.

- Abdullahi, M., et al. (2022). Dialectal equity in crowdsourced speech data. *Proceedings of the ACM Symposium on Computing for Development* (ACM DEV), 1-10. <https://doi.org/10.1145/3563340.3563345>.
- Abdullahi, M., et al. (2022). Gender and dialect bias in African speech AI. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency* (ACM FAccT), 1-12. <https://doi.org/10.1145/3531146.3533217>.
- Adelani, D. I., et al. (2021). Masakhane: Overcoming NLP bias for African languages. *arXiv preprint*. arXiv:2110.07586.
- Adelani, D. I., et al. (2023). Hausa MT: Pitfalls and progress. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics* (EACL), 1-15. <https://doi.org/10.18653/v1/2023.eacl-main.201>.
- Bird, S. (2020). Decolonising speech and language technology. *Proceedings of the 28th International Conference on Computational Linguistics* (COLING), 1-10. <https://doi.org/10.18653/v1/2020.coling-main.493>.
- Financial Times. (2023, October 22). How Africa can gain from artificial intelligence. Retrieved from <https://www.ft.com/content/bdab80fe-e800-4c1c-926d-a6faa750cd57>.
- Garba, S., & Bello, M. (2022). AI in Hausa education: Prospects and challenges. *Nigerian Journal of Technology*, 41 (3), 78-92.
- Haruna, A., Ibrahim, A., & Zubairu, I. (2020). Dialectal variation in Hausa speech recognition systems: A comparative study. *International Journal of Computational Linguistics*, 11 (4), 205-220.
- Independent. (2024, May 2). Kaduna State Government, Google, DSN launch first Hausa-language AI learning series. *Independent Newspaper Nigeria*. Retrieved from <https://independent.ng>.
- Ismaila, A., Yusuf, M., & Zayyanu, U. (2023). Localized chatbots and multilingual learning in West Africa. *Journal of AI for Development*, 3 (1), 10-25.
- Lingvanex. (2024). AI-powered language translation and text-to-speech for Hausa. Retrieved May 10, 2024, from <https://lingvanex.com>.

Mbuk, G. (2021). AI and the preservation of African oral literatures: The case of Hausa proverbs. *Journal of African Languages and AI*, 2 (1), 75-90.

TheCable. (2024, March 18). Kaduna launches initiative to make AI accessible to Hausa speakers. Retrieved from <https://www.thecable.ng/kaduna-launches-initiative-to-make-ai-accessible-to-hausa-speakers>.