

Research Article

## Artificial Intelligent (AI) and Endangered Language Preservation: Challenges Among Igbo, Hausa and Yoruba

Ogbonnaya Princess Onyekachi\*

*Department of Languages, School of general and Basic Studies,*

*Akanu Ibiam Federal Polytechnic, Unwana*

*\*\*Related declarations are provided in the final section of this article*

### ARTICLE INFORMATION

**Received: 20 Jul 2026**

**Accepted: 21 Aug 2026**

**Published: 31 Aug 2026**

**Correspondence Author:**

Ogbonnaya Princess Onyekachi\*

Copyright © 2026 The Author(s). The Authors retain copyright. Licensed under the **Creative Commons Attribution 4.0 International License** ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

Artificial intelligent (AI) is geared recently into a diverse array of innovative tools that empowered linguistic communities to document, preserve, and share endangered languages. Considering the inherent complexities of linguistics, technical limitations, and ethical consideration on human system as an assessing long-term sustainability of language preservation, a focus on AI-driven tools in preserving endangered language rears. This study specifically aims to explore AI preservative potentials in preserving endangered languages specifically the cultural and linguistic subtleties, such as nuances of expression and idiomatic phrases, that define endangered languages among the three major ethnics languages in Nigeria. To systematically draw strings of inferences on that, a cross-sectional research design and quota sampling technique was adopted. 45 married men and women (15 each) from the 3 major ethnic group (Igbo, Hausa, and Yoruba) in Nigeria were sampled using structured questionnaire. They comprised 22 male and 23 females with age ranges from 39 to 67 years ( $M = 51.5$ ,  $SD = 7.17$ ). Descriptive statistics was used for data analysis. The findings identified the widely used AI driven tools across the three tribes in Nigeria; its potential; and technical challenges. It builds on multi-agent Ai frame work and argue that AI can identify, translate, and interpret nuances and tonal subtleties of languages if technological resources, concise computational methods are constituted.

**Keywords:** Artificial Intelligent; Endangered language preservation; Nigeria

## INTRODUCTION

Artificial Intelligence (AI) are leveraged to safeguard and promote indigenous knowledge systems, which often includes endangered languages. The act of leasing language preservation on AI due to its potential may directly or indirectly affect the delicacy and serious risk of losing nuances in communication and social interaction. A suggestion that large language models (LLMs) and computer vision technologies can be used to identify, translate, and interpret low-resource language research assert also that it can provide powerful tools for preservation (Ingram, 2025; Low, et al. 2022; Tella, et al. 2025; Zhong et al., 2024). This suggestion indicates that the applications of large language models (LLMs) using AI in low-resource language research, integrate artificial intelligence with the humanities to preserve and study humanity's linguistic and cultural heritage. Pakray et al. (2025) highlighted that machine learning, deep learning, and cross-lingual transfer learning is a practical technique crucial for preserving linguistic diversity and cultural heritage, indicating a role for AI in preservation. However, little is known about AI capabilities in documenting endangered languages among Igbos, Hausas and Yorubas in Nigeria with recognition of subtle and delicate variation in meaning, tone, feelings that add depth and richness to communication and social interaction. A study by Okonkwo (2025) details how AI's current limitations in interpreting symbolic gestures and proverbs risk flattened the cultural meaning embedded in oral storytelling, particularly in tonal languages like Yoruba or Igbo. He identifies three key trends accelerating the decline of oral histories, including AI's difficulty in processing cultural depth and the general preference for written or visual media over oral transmission. Nzeaka and Ehondor (2024) identifies challenges to language preservation in Nigeria, specifically the difficulty in implementing government action for support, financing, regulations, or language rights legislation. They noted that the introduction of digital platforms featuring dominant foreign languages and minimal national enforcement of language preservation constitute a significant challenge to the transmission of ethnic languages. Their findings suggested that a major obstacle to the use of endangered languages on social media is the absence of digital tools and resources, such as standardized writing systems or digital typefaces for many of these languages. Orji and Korie (2024) explored the current state and prospects of the Igbo language, a major indigenous language spoken in Nigeria, within an AI-driven world and identified challenges faced by the Igbo language in the digital age, where dominant languages and technologies often overshadow lesser-known indigenous languages. Their highlights included that the specific difficulty of developing accurate and reliable Igbo language processing tools is due to limited linguistic resources and data. Akomolede and Olowojebutu (2025) analyzed linguistic complexities and technological limitations, such as tonal variations, diacritic preservation, and data scarcity, hindering Natural Language Processing (NLP) tools for Yorùbá, a major West African language spoken in Nigeria. They identified that AI tools available in Nigeria often lack support for unique linguistic features of Yorùbá, including tonal marks, complex morphologies, and contextual expressions. Exploring key issues related to dangers of leasing language preservation on AI due to its potential should consider the delicacy and serious risk of losing nuances in communication and social interaction.

## PROBLEM OF THE STUDY

There is an unclear understanding about how AI-driven tools could fully capture the cultural and linguistic subtleties, such as nuances of expression and idiomatic phrases, that define endangered languages. Advocates for a community-centered approach to AI development may relatively forget that technology should respect linguistic diversity and preserve cultural authenticity. Pinhanez et al. (2024) narrated how working with indigenous languages creates unique ethical challenges for AI and Natural Language Processing (NLP). Identifying the historical difficulty of developing language technologies for languages due to a lack of resources, linguistic knowledge, and appropriate computational methods for working with small amounts of data would help to understand the unclear situation. These are crucial challenges that could meddle the use of AI in preserving endangered languages. Exploring key ethical issues related to handling language data, capturing linguistic subtleties, and the dangers of releasing data without proper authorizations from language owners, could lead to failure proposal of an alternative preservative agent apart from human system, and could also lead to community-based AI development crises (Perera et al., 2025). Soylu and Sahin (2024) identifies challenges to AI in language preservation, including technical barriers, cultural and social barriers, and financial and resource constraints as being enormous. Detailing the technical barriers such as internet connectivity, device availability, software limitations, and language complexity as hurdles to AI tool effectiveness in remote communities. There is need for multilingual AI policy, ethical governance, and inclusive data architecture to preserve cultural nuance and potentially linguistic diversity, or leaving them for human system preservation. Thus, this perceived need led to the following objectives of the study:

### **Objective of the study**

1. To explore the common AI tool used widely in preserving endangered languages among Igbo, Hausa and Yoruba in Nigeria.
2. To assess AI capabilities in preserving endangered languages among Igbo, Hausa and Yoruba in Nigeria.
3. To explore whether nuances of major languages are genuinely endangered due to digital exclusion.

### **THEORETICAL FRAMEWORK**

A multi-agent AI framework for culturally adaptive translation in underserved language communities proposed by Anik et al. (2025) addressed the failure of existing AI models to capture cultural nuances and idiomatic expressions. This agent specializes on an innovative AI-driven linguistic framework that uses specialized agents for translation, interpretation, content synthesis, and bias evaluation to ensure that both linguistic accuracy and cultural relevance are preserved. Rahman et al. (2025) elaborated the multi-agent AI framework specifically designed for culturally adaptive translation in underserved language communities to counter the marginalization of linguistic diversity by existing efficiency-prioritized AI models. The framework enhances contextual fidelity and mitigates biases through external validation using tools like CrewAI and LangChain, achieving contextually rich and culturally embedded translations for endangered and low-resource languages. Quesada (2025) build an "AI Thinking"

framework, which is meaning-centered and shifts technological development from creating tools for communities to co-creating solutions with them. He articulates a holistic methodology and a five-layer technological ecosystem where communities retain control over their linguistic and cultural knowledge representation. This meaningful solution emerges through the synthesis of cultural understanding, community agency, and technological innovation, moving beyond purely technical approaches to linguistic preservation.

Empirical studies supported the view that AI systems are primarily trained on European and Asian languages, leading to significant underrepresentation and inaccuracies in machine translation of Nigerian languages like Yoruba, Igbo, and Hausa due to algorithmic bias and data scarcity (Ugochukwu, 2025). Tasie (2025) identified the limited availability of digital resources and online content in the endangered Ikwerre language as a key challenge when using AI tools for revitalization. Focusing on the revitalization of the Ikwerre language, an endangered indigenous language in Nigeria, by evaluating the potential and limitations of AI-powered technologies. He argued that the restriction of AI models' ability to process and teach the Ikwerre language is a result of insufficient digital resources. Ogunyemi and Akande (2026) investigated how AI-based tools can support Yoruba language revitalization and promote ethical pedagogy in secondary schools in Ondo State, Nigeria. Their study highlighted that teacher expressed concerns about cultural misrepresentation, tonal inaccuracies, and ethical data use when discussing AI and indigenous language support. Suggesting that the indigenous language of Yoruba remains underrepresented within mainstream AI systems, which increases the risk of further marginalization in digital learning environments. Koc (2025) explores the role of Generative AI (GenAI) and Large Language Models (LLMs) as powerful tools for language preservation, especially for near-native languages, by enabling automated corpus creation, transcription, translation, and tutoring. He identified critical challenges in using AI for endangered language preservation, including data scarcity, ethical issues surrounding ownership and representation, and the need for community-centric data stewardship. Issues related to dangers of leasing language preservation entirely on AI due to its potential should consider the delicacy and serious risk of losing nuances in communication and social interaction. This issue is a question begging for solution despite the level of technological speed in language preservation. These challenges lead to the following research questions:

#### **Research questions:**

1. What is the common AI tool used widely in preserving endangered languages among Igbo, Hausa and Yoruba in Nigeria?
2. What are AI capabilities in preserving endangered languages among Igbo, Hausa and Yoruba in Nigeria?
3. Would AI-drive tools capture nuances of major endangered languages?

#### **METHOD/DESIGN/STATISTICS**

The study adopted cross-sectional research design and quota sampling technique to sample 45 married men and women in academic setting (15 each) from the 3 major ethnic group (Igbo, Hausa, and Yoruba) in Nigeria using structured questionnaire. They comprised 22 male and 23 females with age ranges from 39 to 67 years ( $M = 51.5$ ,  $SD = 7.17$ ). The instrument used was a

structured questionnaire developed by the researcher based on the specific objective of the study. The questions comprised seven structured questions aimed objectively to gather participants' responses, in other to provide a rich understanding of the studied variables through their experiences. This approach ensures that the participants selected can provide detailed insights into the research phenomena. The data was collected and analyzed using modal frequency distribution.

## RESULT

**Table 1:** Modal analysis from three (3) ethnic group in Nigeria and the modal frequency number of responses

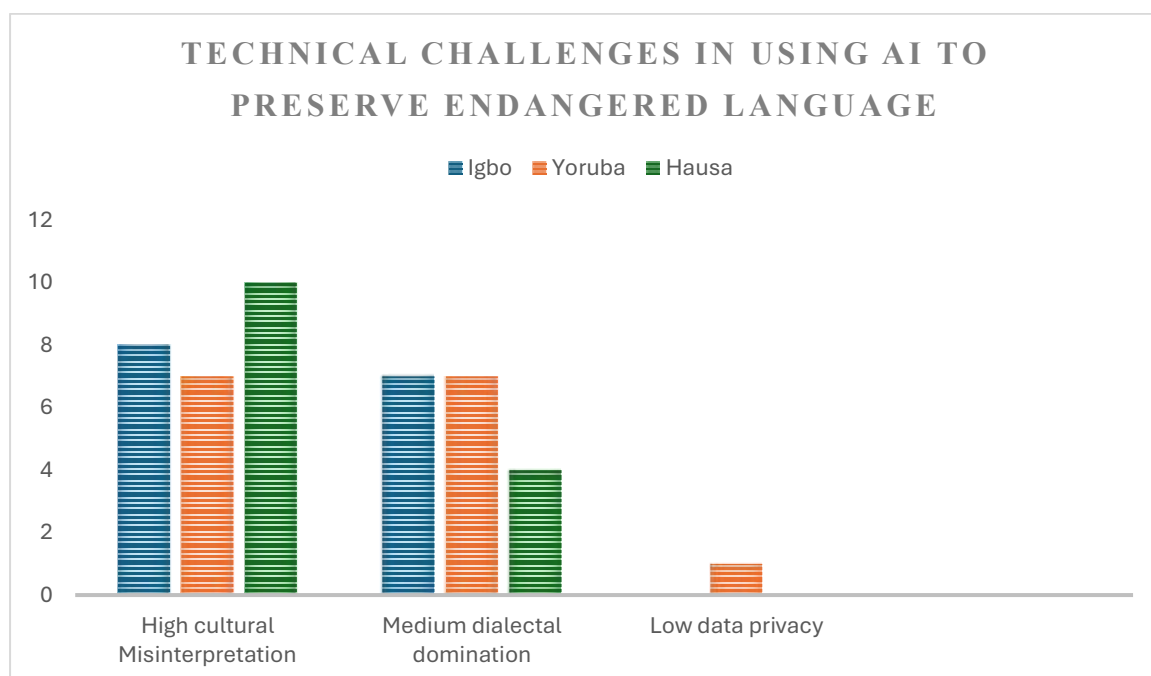
The table above is the modal summary of responses to a particular listed option in each outlined

| S/N | Items                                                                                                       | Option 1                  | Option 2                      | Option 3                |
|-----|-------------------------------------------------------------------------------------------------------------|---------------------------|-------------------------------|-------------------------|
| 1   | Which of the three major Nigerian languages do you work with or speak most fluently?                        | Igbo 15                   | Yoruba 15                     | Hausa 15                |
| 2   | Current AI Tools for language preservation.                                                                 | Generative AI (11)        | Voice recognition machine (9) | Machine translation (7) |
| 3   | How would you rate the accuracy of current AI translation models for endangered languages?                  | Moderate (8)              | Moderate (11)                 | Poor (8)                |
| 4   | In your experience, how well do AI models capture the tonal nuances of Hausa, Yoruba and Igbo?              | Moderate (7)              | Moderate (10)                 | Not moderate (9)        |
| 5   | Do you believe AI will ultimately help or hurt the long-term survival of Igbo, Hausa, and Yoruba languages? | Yes (9)                   | Yes (8)                       | Yes (11)                |
| 6   | To what extent do you feel these major languages are genuinely "endangered" due to digital exclusion?       | Critically endangered (8) | Vulnerable endangered (8)     | Safe (0)                |

question. It stated at the first question the number of participants sampled (15) in each ethnic group. It explores the common AI tool used widely in preserving endangered language; the modal responses gotten includes: Generative AI (11), Voice recognition machine (9), Machine translation (7). The modal responses towards the efficacy of AI in preserving endangered language indicated the following result: Igbo = Moderate (8); Yoruba = Moderate (11); and Hausa

= Poor (8). However, the fourth question that asked how well do AI models capture the tonal nuances of Hausa, Yoruba and Igbo, indicated the following mode: Moderate (7), Moderate (10), and not moderate (9). Conversely, when the participants were asked whether AI will ultimately help or hurt the long-term survival of Igbo, Hausa, and Yoruba languages? Majority replied yes, suggesting significant help as it democratizes access and engages younger generation. When asked the extent they feel nuances of major languages are genuinely “endangered” due to digital exclusion, the response indicate the following: Critically endangered (8), Vulnerable endangered (8), Safe (0). This answered the second research question, further analysis was deployed and it was shown in the below figure 1.

**Figure 1** AI capabilities/challenges in preserving endangered languages in the three different ethnic group in Nigeria.



The figure above shows technical challenges reported in using AI to preserve endangered language across ethnic group. Hausa language owners indicated high cultural misinterpretation of AI tools. Igbo and Yoruba language owners indicated equal medium dialectical domination of standardize dialects over regional variants. Only the Yorubas indicated low data privacy, suggesting unauthorized scraping of cultural folklores, texts, and indigenous knowledge.

## DISCUSSION

This study explored AI preservative potentials in preserving endangered languages specifically the cultural and linguistic subtleties, such as nuances of expression and idiomatic phrases, that define endangered languages among the three major ethnics languages in Nigeria. It answered the three research questions by indicating in each tribe in Nigeria, the common and widely used AI-driven tools in language preservation and exportation which includes: Generative AI (Igbos), Voice recognition machine (Yorubas), Machine translation (Hausas). One among the derived tools supported that of Koc (2025) who discovered the role of Generative AI (GenAI) and Large

Language Models (LLMs) as powerful tools for language preservation, especially for near-native languages, by enabling automated corpus creation, transcription, translation, and tutoring. The second research question was also answered by understanding that AI models capture the tonal nuances of Yoruba and Igbo only and not that of Hausas. The report of the two tribes suggested a significant help of AI-driven tools as it democratizes access and engages younger generation. This report was in line with those of (Ingram, 2025; Low, et al. 2022; Tella, et al. 2025; Zhong et al., 2024) whose suggestion maintained that computer vision technologies (AI-driven tool) can be used to identify, translate, and interpret low-resource language and provide powerful tools for preservation. The third research question was answered indicating that nuances of major languages using AI to preserve them are critically endangered (Igbo), Vulnerable endangered (Yoruba) among the tribes that uses them most. This finding supported that of Ogunyemi and Akande (2026) who found that teacher expressed concerns about cultural misrepresentation, tonal inaccuracies, and ethical data use when discussing AI and indigenous language support. Suggesting that the indigenous language of Yoruba remains underrepresented within mainstream AI systems.

## **IMPLICATION OF FINDINGS**

This study has theoretical and practical implication. Theoretically, the multi-agent AI framework elaborated by Rahman et al (2025) specifically designed for culturally adaptive translation in underserved language communities to counter the marginalization of linguistic diversity by existing efficiency-prioritized AI models have come to stay. The framework enhances contextual fidelity and mitigates biases through external validation using tools like CrewAI and LangChain, achieving contextually rich and culturally embedded translations for endangered and low-resource languages. Practically, the Advocates for a community-centered approach to AI development should work with indigenous languages to create unique ethical solution for AI and Natural Language Processing (NLP). They should identify the historical difficulty of developing language technologies for languages considering lack of resources, linguistic knowledge, and appropriate computational methods for working with small amounts of data would help to preserve endangered languages across the tribe.

## **CONCLUSION**

This study has actually explored the widely used AI preservative tools in Nigeria, its potentials in preserving endangered languages specifically the cultural and linguistic subtleties, such as nuances of expression and idiomatic phrases, that define endangered languages among the three major ethnics languages in Nigeria. The study also discovered subtle challenges using the AI-driven tools for this preservation. It builds on the Multi-agent AI frame work and supported the view that AI-driven tool can be used to identify, translate, and interpret low-resource language and provide powerful tools for preservation. Considering the challenges of AI in interpreting nuances and tonal subtleties of Nigerian languages, the study recommended that Advocates for a community-centered approach to AI development should collaborate with the government to

avail technological resources, linguistic knowledge, and appropriate computational methods for working with large amount of data so to preserve endangered languages across ethnic group.

### **Article Publication Details**

This article is published in the [OpenMind Journal of Multidisciplinary Innovation & Development](#), ISSN: 3108-2920 (Online). In Volume 2 (2026), Issue 4 (Jul - Aug) - 2026

The journal is published and managed by [OMR PUBLICATION](#) .

### **Acknowledgements**

We sincerely thank the editors and the reviewers for their valuable suggestions on this paper.

### **Authors' contributions**

The author read and approved the final manuscript.

### **Funding**

The author declare that no funding was received for this work.

### **Data availability**

No datasets were generated or analyzed during the current study.

### **Declarations**

#### **Ethics approval and consent to participate**

The author declares that not applicable.

#### **Consent for publication**

The author declares that not applicable.

#### **Competing interests**

The author declares no competing interests.

### **REFERENCES**

1. Akomolede, K. K., & Olowojebutu, A. O. (2025). Enhancing digital inclusion through AI-based Yorùbá language localization: Challenges, solutions, and future prospects. *Tech-Sphere Journal for Pure and Applied Sciences*, 2(1).
2. Anik, M. A., Rahman, A., Wasi, A. T., & Ahsan, M. M. (2025, May). Preserving cultural identity with context-aware translation through multi-agent AI systems. In *Proceedings of the 1st Workshop on Language Models for Underserved Communities (LM4UC 2025)* (pp. 51–60).
3. Ingram, M. (2025). The role of AI in language preservation and revitalization. In *AI for community* (pp. 55–80). Chapman and Hall/CRC.
4. Koc, V. (2025). Generative AI and large language models in language preservation: Opportunities and challenges. *arXiv*. <https://arxiv.org/abs/2501.11496>
5. Low, D. S., McNeill, I., & Day, M. J. (2022). Endangered languages: A sociocognitive approach to language death, identity loss, and preservation in the age of artificial intelligence. *Sustainable Multilingualism*, 21(1), 1–25.
6. Nzeaka, E. E., & Ehondor, B. A. (2024). Preservation of indigenous languages, changes in digital media, social development, and family communication. In *Decolonising digital media and indigenisation of participatory epistemologies* (pp. 40–59). Routledge.
7. Ogunyemi, K. O., & Akande, A. A. (2026). AI and Yoruba language revitalization: Ethical development of large language models for pedagogy in Ondo State. *Journal of College of Languages and Communication Arts Education*, 3(1).

8. Okonkwo, O. G. (2025). The erosion of indigenous oral histories in Africa: Globalization and artificial intelligence as dual threats in the digital age. *Trinitarian International Journal of Arts and Humanities*, 1(2).
9. Orji, D. M., & Korie, I. J. (2024). The Igbo language in an AI world: Assessing its current status. *OHAZURUME-Unizik Journal of Culture and Civilization*, 3(1), 137–144.
10. Pakray, P., Gelbukh, A., & Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. *Natural Language Processing*, 31(2), 183–197.
11. Perera, M., Vidanaarachchi, R., Chandrashekeran, S., Kennedy, M., Kennedy, B., & Halgamuge, S. (2025). Indigenous peoples and artificial intelligence: A systematic review and future directions. *Big Data & Society*, 12(2), 20539517251349170.
12. Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., ... Gonçalves, I. (2024). Harnessing the power of artificial intelligence to vitalize endangered indigenous languages: Technologies and experiences. *arXiv*. <https://arxiv.org/abs/2407.12620>
13. Quesada, J. F. (2025). AI thinking as a meaning-centered framework: Reimagining language technologies through community agency. *arXiv*. <https://arxiv.org/abs/2502.14923>
14. Rahman, M. A. A. A., Wasi, A. T., & Ahsan, M. M. (2025). Preserving cultural identity with context-aware translation through multi-agent AI systems. In *Proceedings of the 1st Workshop on Language Models for Underserved Communities (LM4UC 2025)* (p. 51).
15. Soyulu, D., & Sahin, A. (2024). *The role of AI in supporting indigenous languages*.

16. Tasie, C. C. (2025). Artificial intelligence (AI): An innovative tool for the revitalisation of the Ikwerre language. *LALICO Journal of Languages, Literature, and Communication*, 3(2), 113–120.
17. Tella, A., Jatto, E. O., & Ajani, Y. A. (2025). Preserving indigenous knowledge: Leveraging digital technology and artificial intelligence. *IFLA Journal*, 51(3), 703–721.
18. Ugochukwu, L. (2025). The problem of sustainability in translation studies: Whose languages are preserved and whose are erased? *Àkúyííbó Nsugbe Journal of the School of Languages*, 1(3), 200–207.
19. Zhong, T., Yang, Z., Liu, Z., Zhang, R., You, W., Liu, Y., ... Liu, T. (2024). Opportunities and challenges of large language models for low-resource languages in humanities research. *arXiv*. <https://arxiv.org/abs/2412.04497>

## PUBLISHER'S NOTE

OMR PUBLICATION remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of OMR PUBLICATION and/or the editor(s). OMR PUBLICATION disclaims responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.