ARTIFICIAL INTELLIGENCE IN LANGUAGE REVITALIZATION: A SYSTEMATIC REVIEW OF GLOBAL CHALLENGES AND OPPORTUNITIES

Riki Nasrullah
Faculty of Language and Arts, State University of Surabaya
rikinasrullah@unesa.ac.id

ABSTRACT

The rate at which local languages become extinct is alarming and a factor for worry due to globalization and changing cultures. Artificial Intelligence opens opportunities that could support taking care of such languages. The work will seek to conduct a systematic review of the application of Artificial Intelligence in local language revitalization, including AI support for documentation, preservation, and learning, and also pinpoint challenges and opportunities. It adopted the systematic literature review methodology based on data from the Scopus database. Thematic analysis was performed to understand common patterns and opportunities concerning the use of AI. Results showed that, through NLP, ASR, and machine learning, AI could speed up the documentation and learning of local languages, especially in developed countries. However, even in developing countries, the stage is still characterized by some challenges, which involve minimal supporting infrastructure and policies. Conclusion AI can go a long way in supporting local languages, but further investment in technological infrastructure and policies and collaboration between governments, the private sector, and local communities is needed.
Keywords: *Artificial Intelligence, Digital Language Documentation, Endangered Languages, Language Revitalization, Local Languages*.

ABSTRAK

Laju kepunahan bahasa-bahasa lokal sangat mengkhawatirkan dan menjadi faktor yang perlu dikhawatirkan karena globalisasi dan perubahan budaya. Kecerdasan Buatan membuka peluang yang dapat mendukung perawatan bahasa-bahasa tersebut. Penelitian ini akan berusaha melakukan tinjauan sistematis terhadap penerapan Kecerdasan Buatan dalam revitalisasi bahasa lokal, termasuk dukungan AI untuk dokumentasi, pelestarian, dan pembelajaran, serta menunjukkan dengan tepat tantangan dan peluangnya. Penelitian ini menggunakan metodologi tinjauan literatur sistematis berdasarkan data dari basis data Scopus. Analisis tematik dilakukan untuk memahami pola umum dan peluang terkait penggunaan AI. Hasil penelitian menunjukkan bahwa, melalui NLP, ASR, dan pembelajaran mesin, AI dapat mempercepat dokumentasi dan pembelajaran bahasa lokal, terutama di negara-negara maju. Namun, bahkan di negara berkembang, tahap ini masih diwarnai oleh beberapa tantangan, yang melibatkan infrastruktur dan kebijakan pendukung yang minim. Kesimpulan AI dapat sangat membantu dalam mendukung bahasa lokal, tetapi investasi lebih lanjut dalam infrastruktur teknologi dan kebijakan serta kolaborasi antara pemerintah, sektor swasta, dan masyarakat lokal sangat dibutuhkan.
Kata kunci: *Artificial Intelligence, dokumentasi bahasa, Bahasa yang Terancam Punah, Revitalisasi Bahasa, Bahasa Daerah*.

INTRODUCTION

In the digital age and globalization, which has increased the dominance of global languages such as English and Mandarin, local languages are often marginalized (Ajani et al., 2024; Issa et al., 2023; Lee, 2020; Taylor & Kochem, 2022). Young people in most countries would prefer advantageous languages in acquiring social and economic mobility, which are usually national or international languages, rather than local ones (Addaney et al., 2022; Becerra-Lubies et al., 2021; Cassels, 2019; McIvor, 2020; McKenzie, 2022; Whalen et al., 2022). Such a shift translates to the extinction of those local languages much quicker. The speakers of such languages, which have been passed from generation to generation, have become less concentrated. Due to this fact, the revitalization of vernacular languages becomes urgent not to lose completely a part of the cultural heritage in language (Becerra-Lubies et al., 2021; Benyei et al., 2020; L. Chen, 2023; Fujita-Round, 2023; Taylor & Kochem, 2022).

Technology over the last decade has been presenting potential solutions to this challenge. AI has rapidly developed and has contributed much to supporting language preservation efforts in several ways (Arundale, 1999; Pandey et al., 2022; Vo et al., 2024). AI can support language documentation, NLP, speech recognition, and the development of interactive learning applications that would engage the new generation. With projects related to documentation, language training, and automated translation, countries like Canada and New Zealand have already shown how AI can be utilized to protect minority languages (Adnan et al., 2024; Judijanto et al., 2022; Khalid et al., 2023; Taylor & Kochem, 2022;

Younis et al., 2023). These efforts have already shown how AI can serve as an essential tool in the preservation of extinct languages.

However, while AI affords tremendous opportunities to support the preservation of local languages, its utilization is highly uneven worldwide. Several countries have initiated AI-based programs for minority languages. However, in many regions, the use of this technology is still at a relatively early stage and has not been applied in any systematic way. Notably, most AI-language revitalization approaches can't fulfill their prospects due to various technical, social, and political challenges, especially within the developing world where technological infrastructure is minimal, if not nil, with the corresponding political support for the same lacking.

The study will systematically review the application of AI in local language revitalization in various countries. This research aims to synthesize and evaluate multiple studies conducted on how AI has been used to protect and develop minority languages and identify the challenges and opportunities within a global context. The research will also provide a theoretical and practical framework that policymakers and technology developers can apply to support the preservation of local languages in the digital era. It is expected that with this study, there will be a more comprehensive understanding of how minority languages in other countries can be preserved, with solutions that can be offered both locally and globally.

Identifying the use of Artificial Intelligence in the preservation and revival of languages at a local scale, some of the research questions in this regard would aim to explore various aspects related to the use of the said technology in understanding its full potential within the context.

a) How can various artificial intelligence (AI) approaches, such as machine learning, speech recognition, and natural language processing, be used to support the preservation and revitalization of local languages, especially in the context of educational technology and local communities?
b) What are the strategic opportunities and technical, social, and political challenges in implementing AI-based solutions for local language revitalization in different countries, and how can collaboration between government and the private sector play a role in overcoming these barriers?
c) How does integrating AI into the process of documenting, learning, and preserving local languages affect the involvement of indigenous communities, and what are the implications for preserving cultural and linguistic identity in the digital age?

This research is important in two main ways. First, from the point of view of cultural preservation, local languages reflect cultural identity and local knowledge passed down from generation to generation. The loss of local languages means the loss of irreplaceable cultural heritage. Therefore, this research makes an important contribution to finding effective solutions to protect local languages from the threat of extinction in different parts of the world.

From the perspective of technological innovation, this would make AI an innovative new approach to accelerating the revitalization proceedings of minority languages. Because technology happens to be playing an important role in most single aspects of life nowadays, the adoption of AI for the revitalization of local languages is not only an opportunity but an urgent need as well. Therefore, this research is expected to build a premise for policymakers, academics, and practitioners worldwide to develop initiatives based on technology that will preserve local languages.

This research is also important concerning closing the gap in the existing literature; current empirical research on AI use in local language revitalization is still scant. In assessing various studies conducted globally and setting their findings in contexts pertinent to minority language preservation situations in different countries, this research provides new light on how AI can best be utilized in strengthening cultural identity through local language preservation.

METHODS

Research Design

This study employed an aggregative review-based SLR approach. SLR is a systematic and comprehensive way to collect, select, and analyze literature on any topic. An aggregative review approach in SLR aims to address research questions by synthesizing existing research to assess gaps in the study and recommend certain suggestions based on the existing empirical evidence. This research, therefore, goes to the core of

reviewing how AI technology has been applied to revitalize minority languages, especially local languages in Indonesia.

The aggregative reviewing method has been chosen because this approach spells out the method that first prioritizes the collection of similar empirical research findings and then synthesizes them to arrive at strong conclusions or recommendations. Particularly, this is appropriate for assessing the effectiveness of AI in preserving the local language, given that the studies will be borrowed from different geographical and cultural contexts. In contrast, the research design and questions asked are similar. This study, therefore, focuses on review aggregation to present synthesized empirical evidence of common patterns in the use of AI for language revitalization and discuss challenges and opportunities for application to the Indonesian context.

Search Protocol

This study developed the search protocol to ensure broad coverage and high relevance in the literature collection. The only source applied to locate the materials was the Scopus database, among the largest and most credible scientific databases in general, especially for technological and linguistic fields. Scopus was chosen based on its documented ability to provide access to high-quality, empirical research on applying technology, such as AI, in linguistics and language preservation.

The literature search used a combination of keywords formulated using a Boolean strategy. The keywords used were 'artificial intelligence', 'language revitalization', 'language preservation', 'endangered languages', 'local languages', 'minority languages', and 'Indigenous languages'. Here, the Boolean formula is applied: ('Artificial Intelligence' OR 'AI' OR 'Machine Learning') AND ('Language Revitalisation' OR 'Language Preservation' OR 'Endangered Languages') AND ('Local Languages' OR 'Minority Languages' OR 'Indigenous Languages'). This formula allows for a more comprehensive and relevant search by expanding or narrowing the search results according to research needs. The keywords were chosen to cover different aspects of AI technology and its application to preserving regional and minority languages.

Strict inclusion and exclusion criteria were applied to ensure the quality and relevance of the literature. The literature review considered only those articles indexed in Scopus, relevant to AI in the context of language preservation and published within a specified time. These had been designed in a way that would ensure the literature reviewed would meet high academic standards, keeping in mind the objectives of the research.

Inclusion Criteria

1. Studies dealing with the use of AI in the documentation, revitalization, or preservation of minority languages, particularly in relation to regional or endangered languages.
2. Empirical studies using primary or secondary data relevant to the application of AI technology in the context of linguistics, language protection, or language revitalization.
3. Studies using quantitative, qualitative, or mixed methods with a focus on AI applications for minority languages.
4. A review of studies conducted in various global contexts examines the protection of minority languages, with a particular focus on Indonesia. Additionally, an analysis of studies investigating the utilization of AI in the preservation of regional or endangered languages in other countries.
5. The full text of the articles is available and has been subjected to peer review.

Exclusion criteria

1. Studies that do not explicitly discuss the application of AI for language protection or revitalization, for example, studies that only focus on AI in general without a language context or studies that only discuss the technical aspects of AI without its application in language protection.
2. Studies that focus on applying AI outside of linguistic contexts, for example, for general educational purposes, are not directly related to preserving minority or local languages. Articles written in languages other than English, unless a fully accessible translation is available.
3. Articles that are duplicates or already covered in other reviewed articles.
4. Studies that take the form of literature reviews or conceptual discussions that do not contain empirical data.

Synthesis Process

After the literature has been collected through a systematic search process, this research will use thematic analysis as the primary method to synthesize the research results. Thematic analysis is an approach used to identify, analyze, and report patterns (themes) in data. In this context, the thematic analysis will be applied to categorize the findings from various studies based on several main categories:

a) The reviewed studies will be analyzed to find similarities in research design, such as the use of experimental or case study methodologies in evaluating the application of AI for language revitalization.
b) Relevant empirical findings will be grouped based on specific aspects of AI use in language preservation, such as language documentation, automatic translation, and AI-based learning applications.
c) Based on the synthesized results, this research identifies recommendations that various researchers have made regarding the use of AI and how these recommendations can be implemented in the Indonesian context.

In addition, findings were triangulated within the synthesis process to ensure valid and reliable conclusions. This research's use of findings pulled from several sources allows it to give a holistic view of the application of AI in minority language revitalization and identify research gaps that still need to be answered.

RESULTS

The Role of Artificial Intelligence (AI) in Local Language Preservation and Revitalisation

In the face of globalisation and domination, one of the greatest challenges is the preservation of local languages as a means of preserving cultural heritage. In this context, AI can greatly assist in revitalising local languages, especially in the field of machine learning, including speech recognition and natural language processing. In order to overcome resource constraints and to support education and language use in local communities, AI can provide a wide range of applications. (Grönroos et al., 2020; Le et al., 2022).

Machine learning is very advanced in translating and analyzing languages, which is core to maintaining local languages. For instance, models such as the NLLB-200 developed at Meta AI enable the automated translation of low-resource languages, including those deemed to be endangered local languages. In this regard, NLLB-200 successfully enhances the level of access there is to these languages in digital form through the creation of parallel data emanating from several online sources (Buscaldi & Rosso, 2023; Le et al., 2022). However, local dialects and the deficiency in out-of-the-box training based on culture and geography are some of the challenges Putumayo residents face (Le et al., 2022).

In addition, research on the application of Neural Machine Translation to local languages such as Minangkabau demonstrates that when properly optimized, NMT models are capable of producing reasonably good translations even under resource-poor conditions (Grönroos et al., 2020; Mayhew et al., 2020). This should be a line of action taken in a specific manner in relation to characteristics of the local language where its use is targeted for effectiveness.

One area in which AI can help with preserving the local language is ASR. ASR could provide the necessary voice and accent variations of low-resource languages, stemming from various research into Hakka (de Medeiros & Ortiz, 2022; Grönroos et al., 2020; Meyer, 2011). Models like Whisper from Open AI are now being used for digital language learning applications that allow for automatic pronunciation assessment, for example (Grönroos et al., 2020). The above-mentioned technologies will help to document, in digital formats, local languages that were not written so far and also help in the dissemination of those local languages for language education through online platforms or mobile applications. However, the successful implementation of ASR depends upon the availability of adequate and good-quality speech data (Grönroos et al., 2020; Le et al., 2022; Vo et al., 2024)This can also lead to speech recognition failing because of a lack of data to back it up, affecting the effectiveness of the resulting educational applications.

NLP can play an important role in the creation of digital resources for local languages—from online dictionaries to translation tools and text-to-speech systems. Language revitalization programs have put NLP into practice for developing easily accessible text-based learning platforms for the wider community (Grönroos et al., 2020; Le et al., 2022; Selin Akgun & Christine Greenhow, 2021). One of the

major challenges in using NLP with local languages is that there is a shortage of suitable training data. As was underlined in the data-driven language preservation studies, AI models based only on written data cannot capture the dynamics of everyday language use (Grönroos et al., 2020). Thus, it is of critical importance to involve the native-speaking community in the documentation process so the NLP model matches the local cultural context (Le et al., 2022).

Opportunities and Challenges of Artificial Intelligence (AI) Implementation in Local Language Revitalisation

In the era of rising globalization, the need to save local languages from oblivion has become more and more imperative, especially for those who are considered to be at risk. Consequently, regarding promoting or revitalizing such local languages, artificial intelligence has come to the fore as a potential tool for support. However, its implementation requires a strategic approach considering manifold technical, social, and policy challenges. In this respect, hence, the government, private sector, academia, and local communities all have important roles to play in the application and optimization of AI toward language preservation.

AI also offers opportunities in many areas, mainly thanks to NLP technology. Such technologies enable the creation of speech recognition models that can be used to accelerate the digitization of languages. The use of the Whisper model for Hakka, for example, shows the trend of how AI systems can support resource-poor languages so that they can be digitized more quickly (Vo et al., 2024). AI programs, such as in the case of No Language Left Behind by Meta AI, have successfully supported automatic translations in low-resource languages like Genoese (Buscaldi & Rosso, 2023). The successors of machine translation contribute to closing the gaps in the digital divide between the majority and minority languages and extending access to digital technology for communities of local language speakers (Ahda et al., 2024).

Besides that, digitization of local language content also extends to projects such as ABAIR supporting the Gaelic language. ABAIR applies AI in the development of text-to-speech and speech recognition applications. These applications support the inclusion of minority languages in parts of daily life in the digital world (Chasaide et al., 2017; C. Y. Chen et al., 2023; L. Chen, 2023). The abovementioned projects highlight AI's support for language preservation in education and across digital communication.

However, the huge potential AI offers rests on some serious technical challenges. In this regard, the most common limiting factors relate to the availability of data. Often, the local languages lack corpus or organized data that may effectively train the models of AI. This creates a bottleneck regarding the development of the models of AI with great accuracy, as it was found on Hakka and Genoese where non-availability of data hindered the translation and speech recognition capability of AI (Buscaldi & Rosso, 2023; L. Chen, 2023). In addition, some languages have morphological complexity and dialectal variations that make language standardization difficult. For example, the challenge of handling dialectal variations and local morphological structures was seen in the study of Genoese (Buscaldi & Rosso, 2023; Yu et al., 2023).

Other most important issues include social challenges that have to be taken into consideration when applying AI in local language revitalization (Alqahtani et al., 2023; Pinhanez et al., 2023). Without the involvement of the community of native speakers, the development and application of AI technology itself is unthinkable (Ventayol-Boada, 2023). Without community involvement, these technologies can easily become 'parachute science', not relevant to the local context, and not tailored to its needs. It is documented that the native-speaking communities must be directly involved to have the technology created to accommodate the social and cultural context of the language variety (Lam et al., 2015; Ventayol-Boada, 2023). Besides, there is a concern that AI will capture the mere technical aspects of the language, while deeper cultural aspects go unnoticed or are not represented (Camacho & Zevallos, 2020). This, in turn, might cause erosion to the very cultural identity connected with the language.

Other determining variables that will influence the successful application of AI to local language revitalization include policy challenges. More often, governmental policy support alone cannot bear the weight of minority languages, such as issues concerning the Genoese language, which has had development problems because of a lack of a regulatory body that sustains language standardization (Buscaldi & Rosso, 2023; Hakkarainen, 2022). In turn, in the case of the so-called language technologies, development might be accelerated through further cooperation between the government and the private sector. Projects like the Formosa Speech Recognition Challenge demonstrate that multisectoral

collaboration plays a big role in providing much-needed data and infrastructure for language-based AI projects. (Jian et al., 2015; Liao et al., 2020; Zhu, 2020).

These technical, social, and policy challenges must be met with coherent cooperation from various stakeholders: government, academia, the private sector, and the local community. First, the government has a duty to create supportive policies; afterward, it can be supported by the private sector in infrastructural building and offering other necessary facilities. In this way, AI technology can be used to its fullness to support the preservation of local languages and cultures (Bensemann et al., 2023).

The Influence of Artificial Intelligence in Preserving Local Language and Cultural Identity

What has been further taking place in the effort towards the preservation of endangered languages is the use of artificial intelligence in the documentation, learning, and conservation of local languages. AI is both a boon and a bane since this keeps up the question of its influence on the engagement of native-speaking communities and the preservation of their cultural identity. The forthcoming section discusses in detail the effect of integration of AI in the given context.

The Impact of AI on Local Language Documentation and Preservation Processes

AI has indeed shown remarkable potential in making the process of documenting local languages more efficient. The ASR technology utilized in research on Hakka using the Whisper Model manifestly constitutes one such example of how AI can accelerate digitization for low-resource languages. This technology has enabled languages that were once difficult or impossible to document to be recorded in a very efficient and effective manner, which is an essential step toward preserving oral history and daily usage of the languages within the community of their native speakers (Buscaldi & Rosso, 2023; C. Y. Chen et al., 2023; Ventayol-Boada, 2023).

This, in essence, instills a fear that it might be overly dependent upon language data that do not capture the dynamic dimensions of language used in cultural and social contexts. While all AI models may identify grammatical bits and linguistic structures, they mostly fail to show cultural nuances associated with the language. Language is not only one aspect of communication but also very revealing of culture and heritage. It is in this regard that the aspects of language documentation, based on sole linguistic aspects, tend to make the analyst fail to go further in the revelation of deeper cultural values (Pinhanez et al., 2023).

The Impact of AI on Local Language Learning

AI has also been highly involved in recent times in the development of local language learning applications (Low et al., 2022). For example, An app based on TTS, for example, was developed in the ABAIR project for Irish, bringing students' awareness to phonology and helping them learn phonetic features of a language with ease (Chasaide et al., 2017). Thus, this technology opens access for the young generation to learn and further utilize the local languages in the modern digital era, which is important for retaining language continuity.

Besides, AI was a tool employed in testing the facility of the pronunciation and comprehension of local languages. Researchers used the use of ASR for language education in Hakka to describe that this technology can identify the level of students automatically concerning pronunciation, hence helping in effective learning (L. Chen, 2023; Foley et al., 2018; Masuda-Katsuse, 2016). AI, in that respect, has been a preservative tool for creativity and instruction, greatly contributing to widening the use of local languages in larger communities.

Impact on Native Community Engagement

Another very important factor that needs consideration for the integration of AI to preserve a particular local language is the involvement of the native-speaking community (Liu et al., 2022). One of the major challenges involves the non-availability of active participation from main stakeholders while trying to apply this AI technology to native-speaking communities. In fact, Meta AI studies indicate that, sans their contribution, such technologies gradually devolve into parachute science, in which solutions are identified without any concern for local needs and contexts (Alkaeed et al., 2024). Hence, developing technology at each stage requires the involvement of a native-speaking community if languages are ever to be relevant and function according to cultural values.

Besides, there is a fear that applying AI to document the language will lead to the erosion of cultural values associated with the language. Language is not just a means of communication; it also speaks volumes about a community's cultural identity. The relations could be different not considering the

socio-cultural environment while focusing on the linguistic aspect alone, which might result in the erosion of identity typical and dear to the community speaking the original language (Camacho & Zevallos, 2020; Shi et al., 2021). Therefore, documentation or learning of languages through AI needs to be kept in cognizance of the cultural values of the language.

Cultural Identity Preservation in the Digital Age

AI integrated into local language preservation offers opportunities to make the local languages more accessible in the digital world. According to Buscaldi & Rosso (2023), AI is helpful so that endangered languages may become more visible and used by the public and in digital spaces in a project supporting translation for more than 200 low-resource languages, such as in the No Language Left Behind (NLLB-200). In access, local languages can be preserved in an era of globalization dominated by most languages.

However, one of the most significant challenges is retaining local context within the AI applications. For example, a translation of Genoese using NLLB-200 found that it was not quite accurate and did not handle the names of places and other local terms well (Buscaldi & Rosso, 2023). That means losing out on the cultural authenticity in the usage of the language itself. This underlines how, though AI is making access to the local languages easier, its technology must ensure retention of the cultural and social importance of the language it is digitizing (Dwivedi et al., 2020; Lam et al., 2015).

DISCUSSION

Therefore, it is in this present study that AI has been proposed as an applicable tool to support the revitalization of local languages. Generally, the overall findings from this study indicate that through technologies like NLP, ASR, and machine learning, AI can definitely play a significant role in the process of documentation, learning, and preservation of the local language falling under the processes of endangered local languages. These technologies help in faster documentation and digitization of the endangered languages and equally allow the young more interactive learning of the language. While Artificial Intelligence is promising with many opportunities, it equally faces key technical, social, and policy challenges.

The process of local language documentation will accelerate with AI, especially in those areas of spoken languages that used to be extremely hard to transcribe and recognize. Variations of different technologies, say Whisper Model in Hakka, have shown us the capabilities of AI in digitizing such low-resource languages (Alkaeed et al., 2024; Kan et al., 2024; Liao et al., 2023). The second contribution of AI has been to developing more game-oriented, more accessible language learning apps developed to create the TTS for the ABAIR project for Gaelic (Chasaide et al., 2017; Nance, 2021; Taylor & Kochem, 2022). First of all, it is also very crucial in the successful application of these AI technologies that the community native to the language in question be involved. The very same research has also documented that AI-assisted documentation of local languages runs the real risk of losing social and cultural context without active community involvement.Top of Form

Bottom of Form

AI use has been proven to hasten the process of digitization and documentation of local languages. However, while AI is able to grasp the technical features of a language—such as grammar and phonetics—it often misses or fails to capture deeper cultural meanings. A local language is not just a means of communication; it is also a way through which a community expresses itself culturally and reiterates its oral history. Therefore, AI-assisted language preservation should focus more on the broader cultural dimensions to avoid missing out on the deeper meanings.

This finding is supported by several previous studies, which have documented the role of AI in the preservation of minority languages, including the Canadian and New Zealand ones. Indeed, AI applications are found useful in many aspects of the documentation of minority languages in both countries, training of languages, and automated translation, thus serving to preserve endangered languages (Demissei et al., 2024; Li & Brar, 2022; Safdar et al., 2023). The specifics of the Inuktitut language were implemented with such AI technology that allowed it to obtain an extensive, accurate digital corpus and, combined with relevant tech-based learning tools, handed over to local communities (Eysenbach, 2023; Kadir, 2021; Slavina & Khairova, 2016). AI work in New Zealand supports the language in developing automatic speech recognition systems and machine learning models that help preserve the grammar and phonology of the language in digital formats accessible to a younger generation (Ellis et al., 2023; Li & Brar, 2022; Low et al., 2022; McEwan, 2019).

However, it was revealed in this study that AI applications for minority language preservation face more challenges in developing countries because of the main limiting factors: technological infrastructure restricted by an inability to be online without breaks, shortage of supporting hardware, and lack of human resources trained to manage AI technologies (Cadotte et al., 2022; Le et al., 2022). Besides, there are even more serious barriers to a lack of policy support by the local governments in investing in innovative technologies for language preservation. The current study has identified a number of developing countries that failed to provide enough policies that can favor the implementation of AI in the context of preserving minority languages. The latter is usually better positioned than most existing developed countries in terms of better technological infrastructure and strong policy support, sufficient financial, and human resources.

Besides, the gap in AI application between developed and developing countries can be further created by a limitation in the development of training data for minority languages. To this end, Nyugha (2024) and Raiaan et al. (2024) have suggested that the minority languages of developing countries seldom provide an adequate digital corpus or sufficient volume to train an accurate model. This can be further proved by a number of studies on AI uses in Africa; for instance, even languages such as Yoruba and Zulu are not well-represented digitally, and hence it is also hard to develop appropriate AI technologies. On the other hand, projects on AI in developed nations are usually supported by a sizeable structured language corpus that allows the machine learning models to be better trained and therefore make accurate outputs.

Ehimuan et al. (2024) thus reiterated that the findings in this study expose the need for more investment in digital infrastructure development, policy support, and human resource training that developing countries are supposed to invest in. This, they opined, could alone reduce the existing gap in technology and provide equal usability of AI in minority language preservation across most of the world. This would also involve the creation of such an ecosystem with international cooperation in terms of technology transfer and financing by developed countries for developing countries to have an inclusive and sustainable future implementation of AI.

Theoretical and Practical Implications

Theoretically, this research can contribute to the knowledge regarding the potential of AI in preserving the local language at an international level. This will not only be applied as an effective tool for languages' digitalization but also as a new method in order to enhance involvement within native-speaking communities. The pragmatic implication of the research, therefore, would be that there has to be collaboration within government, private sectors, academia, and local communities for proper inclusivity in the implementation of AI to suit community needs. It will also be important to ensure technical challenges are overcome, which include data limitations, and assure potency in AI technologies while preserving cultural meanings within documented local languages.Top of Form

Research Limitations and Suggestions for Future Research

This study is limited in some aspects. First, during the literature search, this study uses only the Scopus database as a source. Although it provides a wide database, Scopus cannot exclude that relevant literatures from another source are reached. Second, this study focuses on reviewing and includes more empirical data related to AI implementation in a number of local contexts. Long-term impacts related to the use of AI during the maintenance of cultural identity in the digital era require further research. In the near future, more empirical studies conducted in diverse countries will be needed for the socio-cultural and policy aspects of such AI applications that take care of local language preservation.

CONCLUSION

This study demonstrates the huge potential of AI in supporting local language revitalization efforts across the world through NLP technologies, automatic speech recognition, and machine learning. With AI, it will be possible to accelerate the documentation rate of endangered languages and make experience opportunities for learning more interactive and accessible. These findings are in line with research conducted on developed countries like Canada and New Zealand that have begun to use AI in minority language protection through the automated documentation of the language, language training, and translation.

While the study also identified grave challenges that are facing the application of AI in developing countries; these things had to do mostly with limitations in technological infrastructure besides serious policy support. Inadequate data along with the minimal digital language corpus poses a

great barrier to develop true models of AI. These are some of the challenges that point to how uneven developed and developing nations have started paying attention to the use of AI technology.

It also requires bridging the gap by increasing investments in the development of digital infrastructure, strengthening policy support, and training human resources. For this, the second challenge calls for collaboration among the government, private sectors, academia, and communities so that the application of AI can be practical yet inclusive in preserving minority languages. International support is urgently sought through technological cooperation, funding, and other means so an ecosystem can be more sustainable.

Theoretically, the study contributes to the literature of the role that AI plays in the preservation of local languages, particularly in the area of cultural identity and the involvement of native-speaking communities. In practical terms, this research puts forward recommendations on how to create a more inclusive professional collaboration framework that will help key stakeholders overcome both technical and policy challenges in the implementation of AI.

There are several limitations to this study. First, Scopus is the only literature source in this study; as such, relevant literature from other databases probably remains reached. Second, this is a literature review study and therefore does not include deeper empirical data on the use of AI in various local contexts. More comprehensive empirical research will be required in future studies, especially for developing countries, to be better positioned to understand how AI can be integrated into local social, cultural, and policy considerations. Future studies shall also discuss some long-term implications for the integration of AI on the sustainability of cultural and linguistic identities in the digital era.

REFERENCES

Addaney, M., Yegblemenawo, S. A. M., Akudugu, J. A., & Kodua, M. A. (2022). Climate change and preservation of minority languages in the upper regions of Ghana: A systematic review. *Chinese Journal of Population, Resources and Environment*, *20*(2), 177–189. https://doi.org/https://doi.org/10.1016/j.cjpre.2022.06.008

Adnan, M., Xiao, B., Ali, M. U., Bibi, S., Yu, H., Xiao, P., Zhao, P., Wang, H., & An, X. (2024). Human inventions and its environmental challenges, especially artificial intelligence: New challenges require new thinking. *Environmental Challenges*, *16*, 100976. https://doi.org/https://doi.org/10.1016/j.envc.2024.100976

Ahda, F. A., Wibawa, A. P., Prasetya, D. D., & Sulistyo, D. A. (2024). Comparison of Adam Optimization and RMSprop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation. *International Journal on Informatics Visualization*, *8*(1), 231–238. https://doi.org/10.62527/joiv.8.1.1818

Ajani, Y. A., Oladokun, B. D., Olarongbe, S. A., Amaechi, M. N., Rabiu, N., & Bashorun, M. T. (2024). Revitalizing Indigenous Knowledge Systems via Digital Media Technologies for Sustainability of Indigenous Languages. *Preservation, Digital Technology and Culture*, *53*(1), 35–44. https://doi.org/10.1515/pdtc-2023-0051

Akgun, S., & greenhow, C. (2021). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *2021 Sep 22*, *2*(3), 431–440. https://doi.org/10.1007/S43681-021-00096-7

Alkaeed, M., Qayyum, A., & Qadir, J. (2024). Privacy preservation in Artificial Intelligence and Extended Reality (AI-XR) metaverses: A survey. *Journal of Network and Computer Applications*, *231*(June), 103989. https://doi.org/https://doi.org/10.1016/j.jnca.2024.103989

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy*, *19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Arundale, R. B. (1999). An alternative model and ideology of communication for an alternative to politeness theory. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, *9*(1), 119–153. https://doi.org/10.1075/prag.9.1.07aru

Becerra-Lubies, R., Mayo, S., & Fones, A. (2021). Revitalization of indigenous languages and cultures: critical review of preschool bilingual educational policies in Chile (2007–2016). *International Journal of Bilingual Education and Bilingualism*, *24*(8), 1147–1162. https://doi.org/10.1080/13670050.2018.1563584

Bensemann, J., Brown, J., Witbrock, M., & Yogarajan, V. (2023). Is it Possible to Preserve a Language

using only Data? *Cognitive Science: A Multidisciplinary Journal*, *47*(6), 1–5. https://doi.org/10.1111/cogs.13300

Benyei, P., Arreola, G., & Reyes-García, V. (2020). Storing and sharing: A review of indigenous and local knowledge conservation initiatives. *Ambio*, *49*(1), 218–230. https://doi.org/10.1007/s13280-019-01153-6

Buscaldi, D., & Rosso, P. (2023). How good is NLLB-200 for low-resource languages? A study on Genoese. *9th Italian Conference on Computational Linguistics*, 1–4.

Cadotte, A., Le, N. T., Boivin, M., & Sadat, F. (2022). Challenges and Perspectives for Innu-Aimun within Indigenous Language Technologies. *5th Workshop on the Use of Computational Methods in the Study of Endangered Languages, Proceedings of the Workshop*, 99–108. https://doi.org/10.18653/v1/2022.computel-1.13

Camacho, L., & Zevallos, R. (2020). Language technology into high schools for revitalization of endangered languages. *27th International Conference on Electronics, Electrical Engineering and Computing, INTERCON 2020*, 2018–2021. https://doi.org/10.1109/INTERCON50315.2020.9220197

Cassels, M. (2019). Indigenous languages in new media: Opportunities and challenges for language revitalization. *WPLC (Working Papers of the Linguistics Circle)*, *29*(1), 25–43. https://journals.uvic.ca/index.php/WPLC/article/view/18767

Chasaide, A. N., Chiaráin, N. N., Wendler, C., & Berthelsen, H. (2017). The ABAIR ini tiative : Bringing Spoken Irish into the Digital Space. *Interspeech*, *1*(1), 2113–2117.

Chen, C. Y., Chang, C. C., & Hsu, Y. H. (2023). Accelerating Hakka Speech Recognition Research and Development Using the Whisper Model. *The 35th Conference on Computational Linguistics and Speech Processing*, *Rocling*, 366–369.

Chen, L. (2023). Assessing Language Vitality and Sustainability of Minor Chinese Dialects: A Case Study of Dapeng, a Hakka–Cantonese Mixed Dialect. *Sustainability*, *15*(9), 1–18. https://doi.org/10.3390/su15097478

de Medeiros, N. M., & Ortiz, K. Z. (2022). Formal language assessment in low-educated persons with aphasia: can the lesion effect be distinguished from the education effect? *Arquivos de Neuro-Psiquiatria*, *80*(2), 125–128. https://doi.org/10.1590/0004-282X-ANP-2020-0475

Demissei, D. B., Biratu, T. D., Gamshe, E. N., & Deressa, A. T. (2024). Attitude towards assisted reproductive technology: acceptance of donors eggs, sperms, and embryos as treatment of human infertility: a systematic review and meta-analysis. *Reproductive Health*, *21*(1), 1–17. https://doi.org/10.1186/s12978-024-01741-0

Dwivedi, P., Shraddha, C., Mathews, S., Majumder, S., Madhumathi, R., & Vasundhara, M. R. (2020). Predicting Language Endangerment: A Machine Learning Approach. *2020 11th International Conference on Computing, Communication and Networking Technologies, ICCCNT 2020*, 1–7. https://doi.org/10.1109/ICCCNT49239.2020.9225576

Ehimuan, B., Anyanwu, A., Olorunsogo, T., Akindote, O. J., Abrahams, T. O., & Reis, O. (2024). Digital inclusion initiatives: Bridging the connectivity gap in Africa and the USA – A review. *International Journal of Science and Research Archive*, *11*(1), 488–501. https://doi.org/10.30574/ijsra.2024.11.1.0061

Ellis, N., Macdonald, E., & Almeida, E. (2023). Taonga in a digital world: Maori adornment and the possibilities of reconnection. *Journal of the Royal Society of New Zealand*, *53*(3), 362–380. https://doi.org/10.1080/03036758.2022.2090967

Eysenbach, G. (2023). The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. *JMIR Medical Education*, *9*. https://doi.org/10.2196/46885

Foley, B., Arnold, J., Coto-Solano, R., Durantin, G., Ellison, T. M., van Esch, D., Heath, S., Kratochvíl, F., Maxwell-Smith, Z., Nash, D., Olsson, O., Richards, M., San, N., Stoakes, H., Thieberger, N., & Wiles, J. (2018). Building Speech Recognition Systems for Language Documentation: The CoEDL Endangered Language Pipeline and Inference System (Elpis). *6th Workshop on Spoken Language Technologies for Under-Resourced Languages, SLTU 2018*, 205–209. https://doi.org/10.21437/SLTU.2018-43

Fujita-Round, S. (2023). Language Revitalization and the Classroom: Video Workshops at an Elementary School in Miyakojima. *Languages*, *8*(1). https://doi.org/10.3390/languages8010004

Grönroos, S. A., Virpioja, S., & Kurimo, M. (2020). Morfessor EM+Prune: Improved subword segmentation with expectation maximization and pruning. *12th International Conference on*

Language Resources and Evaluation, Conference Proceedings, *May*, 3944–3953.

Hakkarainen, H. (2022). The Cultivation of Emotions in the Press: Searching for 'Education of the Heart' in German-Language Digital Newspaper Collections. *Media History*, *28*(4), 465–481. https://doi.org/10.1080/13688804.2021.2013182

Issa, M., Khan, H. K., Hussain, S., & Ali, A. (2023). Language erosion: an overview of declining status of indigenous languages of Gilgit-Baltistan, Pakistan. *International Journal of Multicultural Education*, *25*(2), 817–828.

Jian, M.-S., Chen, Y.-C., & Chen, J.-L. (2015). RPG game learning for language education with voice recognition in cloud. In F. S. & M.-P. J. (Eds.), *Frontiers in Artificial Intelligence and Applications* (Vol. 275, pp. 173–183). IOS Press. https://doi.org/10.3233/978-1-61499-503-6-173

Judijanto, L., Asfahani, A., Prusty, A., Krisnawati, N., & Bakri, A. A. (2022). Industrial Revitalization with AI between Opportunities and Challenges for Global Economic Growth. *Journal of Artificial Intelligence and Development*, *1*(1), 49–57. https://edujavare.com/index.php/JAI/article/view/340

Kadir, R. (2021). Language maintenance in an interracial marriage: The case of Indonesian females' language choice in Canada. *Indonesian Journal of Applied Linguistics*, *11*(1), 146–156. https://doi.org/10.17509/ijal.v11i1.34589

Kan, P. F., Jones, M., Meyers-Denman, C., & Sparks, N. (2024). Emergent bilingual children during the silent period: A scoping review of their communication strategies and classroom environments. *Journal of Child Language*, 1–34. https://doi.org/DOI: 10.1017/S0305000924000151

Khalid, N., Qayyum, A., Bilal, M., Al-Fuqaha, A., & Qadir, J. (2023). Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Computers in Biology and Medicine*, *158*, 106848. https://doi.org/https://doi.org/10.1016/j.compbiomed.2023.106848

Lam, K. N., Al Tarouti, F., & Kalita, J. (2015). Automatically creating a large number of new bilingual dictionaries. *The National Conference on Artificial Intelligence*, *3*, 2174–2180. https://doi.org/10.1609/aaai.v29i1.9493

Le, N. T., Cadotte, A., Boivin, M., & Sadat, F. (2022). Deep Learning-Based Morphological Segmentation for Indigenous Languages: A Study Case on Innu-Aimun. *3rd Workshop on Deep Learning Approaches for Low-Resource NLP, Proceedings of the DeepLo Workshop*, 146–151. https://doi.org/10.18653/v1/2022.deeplo-1.16

Lee, N. H. (2020). The Status of Endangered Contact Languages of the World. *Annual Review of Linguistics*, *6*(Volume 6, 2020), 301–318. https://doi.org/https://doi.org/10.1146/annurev-linguistics-011619-030427

Li, J., & Brar, A. (2022). The use and impact of digital technologies for and on the mental health and wellbeing of Indigenous people: A systematic review of empirical studies. *Computers in Human Behavior*, *126*, 106988. https://doi.org/https://doi.org/10.1016/j.chb.2021.106988

Liao, Y.-F., Chang, C.-Y., Tiun, H.-K., Su, H.-L., Khoo, H.-L., Tsay, J. S., Tan, L.-K., Kang, P., Thiann, T.-G., Iunn, U.-G., Yang, J.-H., & Liang, C.-N. (2020). Formosa Speech Recognition Challenge 2020 and Taiwanese across Taiwan Corpus. In *23rd Conference of the Oriental COCOSDA* (pp. 65–70). https://doi.org/10.1109/O-COCOSDA50338.2020.9295019

Liao, Y.-F., Hwang, S.-H., Chen, Y.-S., Lai, H.-C., Chung, Y.-H., Shen, L.-T., Huang, Y.-C., Huang, C.-J., Han, H. W., Chen, L.-W., Su, P.-C., & Huang, C.-S. (2023). Taiwanese Hakka Across Taiwan Corpus and Formosa Speech Recognition Challenge 2023 - Hakka ASR. *2023 26th Conference of the Oriental COCOSDA International Committee for the Co-Ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA)*, 1–6. https://doi.org/10.1109/O-COCOSDA60357.2023.10482979

Liu, Z., Richardson, C., Hatcher, R., & Prud'Hommeaux, E. (2022). Not always about you: Prioritizing community needs when developing endangered language technology. *Annual Meeting of the Association for Computational Linguistics*, *1*, 3933–3944. https://doi.org/10.18653/v1/2022.acl-long.272

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. https://doi.org/10.2478/sm-2022-0011

Masuda-Katsuse, I. (2016). Support system for pronunciation instruction and practice in special education classes for language-disabled children. *Proceedings of Meetings on Acoustics*, *29*(1). https://doi.org/10.1121/2.0000498

Mayhew, S., Bicknell, K., Brust, C., McDowell, B., Monroe, W., & Settles, B. (2020). Simultaneous translation and paraphrase for language education. *Proceedings of the Annual Meeting of the*

Association for Computational Linguistics, 232–243. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85115686939&partnerID=40&md5=4e97ea6d42dfd7040652dcf140904b97

McEwan, R. (2019). Iwi radio in the era of media convergence: The opportunities and challenges of becoming 'more than radio.' *Pacific Journalism Review*, *25*(1–2), 139–157. https://doi.org/10.24135/pjr.v25i1.469

McIvor, O. (2020). Indigenous Language Revitalization and Applied Linguistics: Parallel Histories, Shared Futures? *Annual Review of Applied Linguistics*, *40*, 78–96. https://doi.org/DOI: 10.1017/S0267190520000094

McKenzie, J. (2022). Addressing historical trauma and healing in Indigenous language cultivation and revitalization. *Annual Review of Applied Linguistics*, *42*, 71–77. https://doi.org/10.1017/S0267190521000167

Meyer, B. (2011). Effect of speech-intrinsic variations on human and automatic recognition of spoken phonemes. *Journal of the Acoustical Society of America*, *129*(1), 388–403. https://doi.org/10.1121/1.3514525

Nance, C. (2021). Scottish Gaelic revitalisation: Progress and aspiration. *Journal of Sociolinguistics*, *25*(4), 617–627. https://doi.org/10.1111/josl.12508

Nyugha, P. G. (2024). The Importance of Artificial Intelligence to Africa's Devel-opment Process: Prospects and Challenges. *PriMera Scientific Engineering*, *4*(6), 20–30.

Pandey, A., Singh, A., Singh, G., & Rakhra, M. (2022). Using AI and IoT to assess the efficacy of English-language curricula in higher education: A Proposed Method. *AIST 2022 - 4th International Conference on Artificial Intelligence and Speech Technology*. https://doi.org/10.1109/AIST55798.2022.10065284

Pinhanez, C. S., Cavalin, P., Vasconcelos, M., & Nogima, J. (2023). Balancing Social Impact, Opportunities, and Ethical Constraints of Using AI in the Documentation and Vitalization of Indigenous Languages. *IJCAI International Joint Conference on Artificial Intelligence*, 6174–6182. https://doi.org/10.24963/ijcai.2023/685

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., Ahmad, J., Ali, M. E., & Azam, S. (2024). A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access*, *12*, 26839–26874. https://doi.org/10.1109/ACCESS.2024.3365742

Safdar, S., Chahar Mahali, S., & Scott, C. (2023). A critical review of multiculturalism and interculturalism as integration frameworks: The case of Canada. *International Journal of Intercultural Relations*, *93*, 101756. https://doi.org/https://doi.org/10.1016/j.ijintrel.2023.101756

Shi, J., Amith, J. D., Chang, X., Dalmia, S., Yan, B., & Watanabe, S. (2021). Highland Puebla Nahuatl–Spanish Speech Translation Corpus for Endangered Language Documentation. *1st Workshop on Natural Language Processing for Indigenous Languages of the Americas, AmericasNLP 2021*, 53–63. https://doi.org/10.18653/v1/2021.americasnlp-1.7

Slavina, L. R., & Khairova, I. V. (2016). Language education in the regions of Russia and Canada: A comparative analysis. *Mathematics Education*, *11*(6), 1809–1816. https://www.scopus.com/inward/record.uri?eid=2-s2.0-84989863848&partnerID=40&md5=c24e603ec3b3ff93acd39169b06007c2

Taylor, J., & Kochem, T. (2022). Access and empowerment in digital language learning, maintenance, and revival: a critical literature review. *Diaspora, Indigenous, and Minority Education*, *16*(4), 234–245. https://doi.org/10.1080/15595692.2020.1765769

Ventayol-Boada, A. (2023). Applications of classification trees for endangered language description: Finite verb morphology in Kolyma Yukaghir. *6th Workshop on the Use of Computational Methods in the Study of Endangered Languages, Proceedings of the Workshop*, 7–12.

Vo, H. N. K., Le, D. D., Phan, T. M. D., Nguyen, T. S., Pham, Q. N., Tran, N. O., Nguyen, Q. D., Vo, T. M. H., & Quan, T. (2024). Revitalizing Bahnaric Language through Neural Machine Translation: Challenges, Strategies, and Promising Outcomes. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23360–23368. https://doi.org/10.1609/aaai.v38i21.30385

Whalen, D. H., Lewis, M. E., Gillson, S., McBeath, B., Alexander, B., & Nyhan, K. (2022). Health effects of Indigenous language use and revitalization: a realist review. *International Journal for Equity in Health*, *21*(1), 1–14. https://doi.org/10.1186/s12939-022-01782-6

Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023). A Systematic Literature Review on the Applications of Robots and Natural Language Processing in Education.

Electronics (Switzerland), *12*(13). https://doi.org/10.3390/electronics12132864
Yu, C., Yu, J., Qian, Z., & Tan, Y. (2023). Improvement of Acoustic Models Fused with Lip Visual Information for Low-Resource Speech. *Sensors*, *23*(4), 1–19. https://doi.org/10.3390/s23042071
Zhu, A. (2020). Application of AI Identification Technology in Foreign Language Education. *Proceedings - 2020 International Conference on Artificial Intelligence and Education, ICAIE 2020*, 71–75. https://doi.org/10.1109/ICAIE50891.2020.00024

CURRICULUM VITAE

Complete Name : Riki Nasrullah
Institution : Faculty of Language and Arts, State University of Surabaya
Education : S-3
Research Interests: Neurolinguistik, Dialektologi, Revitalisasi Bahasa Daerah