

The Impact of AI on Endangered Languages: Can Technology Save or Kill?

Shakhnoza Sharofova

BA student, Uzbekistan State University of World Languages
shakh1.1607@gmail.com, +99899 595 80 16

Abstract: This article examines the impact of artificial intelligence (AI) on endangered languages, assessing its potential to contribute to preservation or inadvertently accelerate decline. Beginning with a brief overview of the cultural significance of endangered languages, the article outlines challenges in preservation. It surveys historical preservation methods and introduces AI's role. Positive impacts of AI, such as documentation and revitalization, are discussed alongside concerns, including loss of authenticity. Through case studies, the article navigates successes and challenges. The central question, "Can Technology Save or Kill Endangered Languages?" is explored, emphasizing the delicate balance between technological aid and unintended consequences. The conclusion provides insights into the future of AI in language preservation and recommends responsible implementation. This article contributes to understanding the nuanced relationship between technology and endangered languages in academic discourse.

Keywords: endangered languages, artificial intelligence, language preservation, linguistic diversity, cultural heritage, documentation methods, language revitalization, technological impact, ethical considerations, linguistic authenticity, cultural appropriation, case studies, language learning apps, globalization effects.

Introduction: In the rich tapestry of global cultures, the urgency of preserving endangered languages has never been more poignant. As communities navigate the challenges presented by modernity, the preservation of linguistic diversity stands at a critical juncture. Historically, conventional methods anchored preservation efforts, but the contemporary landscape introduces a novel player—artificial intelligence (AI). In the unfolding narrative of linguistic conservation, the novelty of AI technology offers unprecedented tools for documentation, analysis, and revitalization of endangered languages. This development aligns seamlessly with the topicality of our times, where the convergence of technology and cultural preservation is a subject of both academic and societal discourse.

The actuality of this inquiry lies in the pressing need to explore the potential of AI in mitigating the threats faced by endangered languages in our interconnected world. As globalization and cultural assimilation loom large, the article delves into the present-day relevance of AI as a double-edged sword—potentially serving as a powerful ally in preservation or unwittingly accelerating linguistic decline. Can technology, specifically AI, be the harbinger of linguistic rejuvenation, or does it pose unforeseen risks? This question forms the nucleus of our exploration into the dynamic interplay between technology and the timeless essence of endangered languages

Endangered languages

Endangered languages refer to languages at risk of extinction due to a declining number of speakers. The UNESCO Atlas of the World's Languages in Danger categorizes endangered languages into different levels, ranging from "vulnerable" to "critically endangered" and "extinct." The designation considers factors such as the number of speakers, intergenerational transmission, and documentation. Endangered languages hold immense cultural and linguistic value. They serve as repositories of unique knowledge, reflecting the histories, traditions, and worldviews of the communities that speak them. Linguists argue that each language is a distinct way of understanding and interpreting the world, offering diverse perspectives on human experience. The loss of an endangered language implies not only the disappearance of a means of communication but also the erasure of cultural identity, traditional practices, and oral traditions.

Factors Leading to Endangerment:

1. Globalization: The global spread of dominant languages, often driven by economic and political forces, marginalizes smaller languages. Globalization facilitates the dominance of major languages like English, leading to the gradual erosion of linguistic diversity. Economic opportunities, international trade, and the global influence of media contribute to the adoption of dominant languages, sidelining indigenous and minority languages.

2. Cultural Assimilation: Cultural assimilation occurs when communities shift away from their native languages in favor of more widely spoken languages. Factors such as urbanization, migration, and intermarriage can accelerate this process. The pressure to assimilate into larger cultural and linguistic groups can lead to a decline in the use of indigenous languages, particularly among younger generations.

3. Lack of Documentation: Many endangered languages lack comprehensive documentation, including written records or grammatical descriptions. This absence of documentation hinders efforts to preserve and revitalize languages. Without proper documentation, the linguistic richness and cultural nuances embedded in these languages are at risk of being lost. Documentation is crucial for linguists, educators, and community members working to sustain and revive endangered languages.

4. Language Policies and Education: Inadequate support for minority languages in official language policies and educational systems can contribute to language endangerment. The lack of recognition in formal settings leads to a decline in intergenerational transmission. Without institutional support, speakers of endangered languages may perceive their languages as less valuable or practical, further reducing their use in education and public domains.

5. Erosion of Traditional Practices: Changes in lifestyle, economic structures, and societal norms can contribute to the erosion of traditional practices that involve the use of indigenous languages. Language is deeply embedded in cultural practices, rituals, and ceremonies. As these traditions diminish, the contexts for using the language decrease, accelerating the decline of the language itself.

The Role of Technology in Language Preservation

Oral Tradition and Storytelling: In many cultures, language preservation has been maintained through oral traditions, where stories, myths, and histories are passed down from generation to generation. While effective for cultural transmission, oral methods may lack the permanence and standardization necessary for long-term language documentation. What about endangered languages from cultures with oral traditions? These languages often lack a standardized script and manpower to carry out the painstaking work of transcribing and translating speech.¹ To solve this problem, computer scientists are automating the transcription of endangered languages from audio and aligning audio recordings with their translations into more common languages, helping to preserve these languages for future generations and bypassing the so-called "transcription bottleneck" that has plagued linguistic fieldwork for years.² And language communities which no longer have fluent speakers may one day benefit from advances in automated language reconstruction.

Written Documentation: Writing systems and documentation efforts have been pivotal in preserving languages. This includes dictionaries, grammatical descriptions, and religious texts. Written documentation provides a tangible record of the language but may not capture all aspects, especially oral nuances and pronunciation.

Community-Based Language Programs: Community-driven language programs involve language classes, immersion programs, and cultural events to foster intergenerational transmission. These programs empower communities to take an active role in language preservation but may face challenges in terms of resources and scalability. Humans use language for cooperation, self-expression, and much more. So it's not

¹ Millour, A., & Fort, K. (2020). Text Corpora and the Challenge of Newly Written Languages. Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), 111–120. <https://aclanthology.org/2020.sltu-1.15>

² Zanon Boito, M., Villavicencio, A., & Besacier, L. (2020). Investigating Language Impact in Bilingual Approaches for Computational Language Documentation. Proceedings of the 1st Joint Workshop on Spoken Language Technologies for Under-Resourced Languages (SLTU) and Collaboration and Computing for Under-Resourced Languages (CCURL), 79–87. <https://aclanthology.org/2020.sltu-1.11>

surprising that, since the dawn of computing, scientists have envisioned building technology that can communicate with us seamlessly in natural language—i.e., human (as opposed to programming) language.³ The transition to digital formats enables the creation of multimedia archives, including audio and video recordings, preserving spoken language alongside visual and cultural contexts. Digital archiving allows for more extensive storage and easier access to language resources, facilitating linguistic research and language documentation. Language learning applications and software provide interactive platforms for learners to engage with vocabulary, grammar, and pronunciation. Technology facilitates broader access to language learning resources, making language education more flexible and adaptable to individual needs. ASR technology can transcribe spoken language into written text, aiding in the documentation of oral traditions and spoken narratives. For instance, ASR has been used in projects to transcribe and analyze recordings of endangered languages, making them more accessible for linguistic research. AI-driven machine translation assists in translating texts from one language to another, enabling the translation of written materials for broader audiences. Examples include the use of AI-based translation services to translate written works in endangered languages, making them available to a wider audience and aiding in cross-cultural communication. Language AI tools assist linguists in creating comprehensive documentation of endangered languages, including grammatical analysis and vocabulary compilation. For example, AI-assisted tools for linguists, such as automatic glossing and morphological analysis, streamline the process of creating linguistic resources for endangered languages.

Natural Language Processing (NLP) for Documentation: NLP technologies analyze and extract linguistic patterns from large datasets, aiding linguists in understanding language structures. NLP has been applied to analyze extensive linguistic corpora, helping linguists identify patterns, track language evolution, and inform revitalization strategies. In conclusion, the integration of technology, especially AI, has significantly enhanced language preservation efforts by providing tools for documentation, analysis, and revitalization. These technological advancements play a crucial role in addressing the challenges faced by endangered languages, offering scalable and innovative solutions for linguistic revitalization.

Positive Impacts of AI on Endangered Languages

The advent of Artificial Intelligence (AI) has ushered in a transformative era for the preservation and revitalization of endangered languages, offering multifaceted advantages across linguistic documentation, analysis, and language learning initiatives. In the domain of documentation and linguistic analysis, AI technologies play a pivotal role in automating complex processes, thereby contributing significantly to the preservation of linguistic heritage. Automated transcription and translation tools, powered by AI algorithms, have proven instrumental in accurately transcribing and translating spoken language, enabling the preservation of oral traditions, grammatical structures, and vocabulary nuances inherent in endangered languages. Natural Language Processing (NLP) tools, another facet of AI, provide linguists with powerful analytical capabilities. These tools are adept at discerning linguistic patterns, grammatical structures, and contextual nuances within large datasets. The application of AI-driven morphological analysis further enhances linguistic documentation by unraveling the intricate structures of words and their meanings, ultimately aiding in the creation of dictionaries and comprehensive linguistic resources. Moreover, AI contributes significantly to digital archiving initiatives, facilitating the creation of multimedia archives enriched with audio and video recordings. This not only preserves spoken language but also ensures the accessibility and longevity of linguistic resources. The digital archiving capabilities of AI have proven invaluable in safeguarding cultural heritage embedded in endangered languages, allowing for the preservation of traditional narratives, songs, and rituals. In the arena of language learning, AI-driven applications have emerged as dynamic and personalized tools for learners and speakers of endangered languages.⁴ These applications offer tailored learning experiences by adapting content and difficulty levels based on individual progress, thereby accommodating diverse learning styles and preferences. The integration of gamification elements, such as interactive content, quizzes, and language games, transforms

³ Schubert, L. (2014). Computational Linguistics. In Stanford Encyclopedia of Philosophy. Stanford Center for the Study of Language and Information. <https://plato.stanford.edu/archives/spr2020/entries/computational-linguistics/>

⁴ Languages | UNESCO WAL. (n.d.). En.wal.unesco.org. <https://en.wal.unesco.org/languages>

language learning into an engaging and enjoyable experience, fostering sustained interest and commitment among learners. AI's role in pronunciation assistance cannot be understated. Through sophisticated speech recognition technologies, learners receive real-time feedback on their pronunciation, contributing to the accurate acquisition of phonetics specific to endangered languages. Furthermore, AI promotes accessibility and inclusivity in language learning through features like text-to-speech, making the learning process more approachable for a broader audience. Beyond individual learning experiences, AI-supported language learning apps serve as community-building platforms. Interactive features, discussion forums, and collaborative learning spaces within these applications foster a sense of community among learners and speakers of endangered languages. This communal engagement not only enhances the learning experience but also contributes to the broader revitalization efforts by creating interconnected networks of language enthusiasts.

The positive impacts of AI on endangered languages are profound and wide-ranging, touching upon linguistic documentation, analysis, and language learning. As technology continues to advance, the integration of AI promises to be a linchpin in the collective efforts to preserve and revitalize endangered languages, ensuring their longevity and cultural richness for generations to come.

Criticisms and Concerns:

One prominent concern surrounding the use of AI in language preservation is the potential loss of authenticity in linguistic and cultural heritage. Critics argue that the automated nature of certain AI processes, such as machine translation and transcription, may inadvertently lead to inaccuracies and subtle distortions in the representation of the language. There is a risk that AI, while efficient, may not fully capture the nuances, idioms, and cultural context that are integral to the authentic expression of an endangered language. In this way, the overreliance on technology may compromise the fidelity of language documentation and inadvertently contribute to a standardized and potentially diluted version of the language.

The ethical dimension of using technology in language preservation raises concerns related to cultural appropriation. Critics argue that the development and implementation of AI tools often involve individuals and organizations external to the linguistic community. This raises questions about who controls the technology and its outputs. The potential for outsiders to determine the direction and objectives of language preservation efforts may lead to unintended consequences, including the commodification of linguistic and cultural elements. Additionally, the use of AI-driven language learning apps by individuals outside the endangered language community may raise questions about the respectful engagement with cultural heritage and whether the technology is being utilized in ways that align with the values and wishes of the community. In addressing these concerns, it becomes imperative for developers, linguists, and communities to adopt an ethical and collaborative approach to the integration of AI in language preservation. Efforts should be made to involve and empower local communities in decision-making processes, ensuring that the technology is employed in a manner that respects the authenticity and cultural richness of the endangered language. Moreover, there should be ongoing dialogues between technology developers and linguists to refine AI applications based on feedback from native speakers, thus mitigating the risk of unintended inaccuracies or distortions in language representation. Ultimately, while AI holds significant promise for language preservation, it is crucial to navigate these criticisms and concerns thoughtfully, with a commitment to cultural sensitivity, community engagement, and ethical considerations. The goal should be to leverage technology in a manner that enhances, rather than diminishes, the authenticity and cultural integrity of endangered languages.

Highlight Success Stories:

Maori Language Revitalization in New Zealand:

- Success Story: AI-driven language learning applications have played a significant role in the revitalization of the Maori language in New Zealand. These applications incorporate interactive lessons, gamified elements, and personalized learning experiences, contributing to increased engagement and proficiency among learners. The incorporation of Maori cultural context in AI tools ensures authenticity and alignment with the community's linguistic goals.

Cherokee Language Documentation with AI Transcription:

- Success Story: AI-powered transcription tools have been successfully employed in documenting the Cherokee language. By automating the transcription process, linguists and community members can efficiently convert oral narratives into written form, preserving grammar, vocabulary, and cultural nuances. This has facilitated the creation of digital archives and linguistic resources for the Cherokee language.

ChatGPT for Indigenous Language Interaction:

- Success Story: OpenAI's ChatGPT has been utilized experimentally to interact in indigenous languages. While not specifically designed for language preservation, this demonstrates the adaptability of AI tools to engage with speakers of endangered languages. Such applications can foster dialogue, encourage language use, and contribute to the broader revitalization efforts.

Drawbacks and Failures:

1. Inadequate Representation in Training Data:

- Challenge: AI systems often rely on vast datasets for training. In some cases, the training data may lack sufficient representation of dialectal variations, unique linguistic features, or cultural context specific to endangered languages. This can lead to AI models that struggle to accurately capture the richness of the language, particularly in its cultural and contextual dimensions.

2. Cultural Sensitivity Issues in Machine Translation:

- Challenge: Machine translation tools, while beneficial, may face challenges in accurately translating culturally sensitive terms and idioms. AI algorithms may not fully grasp the intricate cultural nuances embedded in endangered languages, leading to potential misinterpretations and unintentional distortions in translated content.

3. Limited Community Involvement in Tool Development:

- Challenge: Some AI language preservation efforts have faced setbacks due to limited community involvement in the development process. When technology is introduced without active collaboration and input from native speakers and cultural experts, it may not align with the community's goals, resulting in a lack of acceptance and adoption.

4. Accessibility Barriers and the Digital Divide:

- Challenge: The digital divide poses a significant challenge, as not all communities have equal access to technology. In instances where AI-driven language preservation efforts heavily rely on digital platforms, there is a risk of excluding communities with limited technological infrastructure, hindering the inclusivity and reach of such initiatives. Understanding both success stories and challenges is crucial for the continued improvement and ethical deployment of AI in language preservation. By addressing drawbacks and learning from failures, stakeholders can refine strategies, foster community engagement, and enhance the overall effectiveness of AI-driven initiatives in safeguarding endangered languages.

Balancing Act

The question of whether technology can save or potentially accelerate the decline of endangered languages is a complex and delicate balancing act. On one hand, technology offers unprecedented opportunities for language documentation, analysis, and revitalization. Automated transcription, machine translation, and digital archiving tools can significantly contribute to the preservation of linguistic heritage, providing efficient means of documenting grammar, vocabulary, and oral traditions. However, there is a risk that overreliance on technology may inadvertently accelerate the decline of endangered languages. The introduction of global languages through digital platforms, coupled with the allure of economic and educational opportunities associated with major languages, may contribute to language shift and a decrease in intergenerational transmission. If technology is not carefully integrated with community-driven language preservation efforts, there is a potential for it to inadvertently erode the social and cultural contexts that sustain endangered languages. In this delicate balance, it becomes crucial to adopt a holistic approach that combines the benefits of technology with community engagement and cultural sensitivity. Technology should be viewed as a tool to empower communities rather than a substitute for grassroots efforts. Collaborative initiatives that involve linguists, technology developers, and community members can help strike the right balance, ensuring that technology enhances rather than hinders language preservation.

Ethical Considerations:

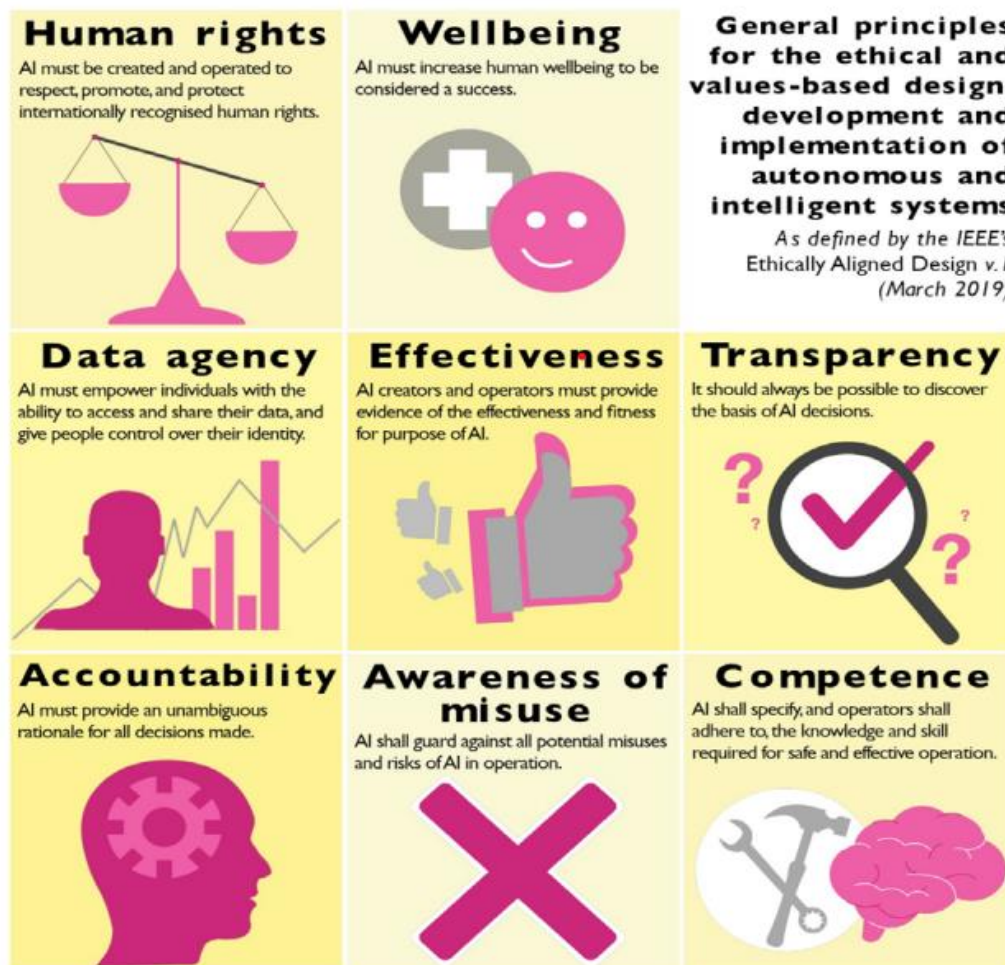
The use of AI in language preservation raises significant ethical considerations that must be carefully navigated. One primary concern is related to cultural appropriation, as the development and application of

AI technologies often involve external entities. Ethical questions arise regarding who controls the technology, the potential for unintended consequences, and whether outsiders may inadvertently dictate the trajectory of language preservation efforts.

Another ethical consideration involves the representation and accuracy of endangered languages in AI models. If training data lacks diverse linguistic features or is not representative of dialectal variations, AI systems may inadvertently perpetuate biases and inaccuracies, compromising the authenticity of language preservation efforts. Moreover, the digital divide poses ethical challenges, as technology-driven language preservation efforts may exclude communities with limited access to digital resources. This raises questions about the inclusivity of AI initiatives and whether they genuinely benefit all speakers of endangered languages.

According to the Foundation for Responsible Robotics, AI must be ethically developed with human rights in mind to achieve their goal of 'responsible robotics', which relies upon proactive innovation to uphold societal values like safety, security, privacy, and well-being. The Foundation engages with policymakers, organises and hosts events, publishes consultation documents to educate policymakers and the public, and creates public-private collaborations to bridge the gap between industry and consumers, to create greater transparency. It calls for ethical decision-making right from the research and development phase, greater consumer education, and responsible law- and policymaking – made before AI is released and put into use.

Figure 2: General principles for the ethical and values-based design, development, and implementation of autonomous and intelligent systems (as defined by the IEEE's Ethically Aligned Design First Edition March 2019)



Conclusion:

In conclusion, the discourse surrounding the integration of technology, particularly Artificial Intelligence (AI), in the preservation of endangered languages necessitates a nuanced examination of its potential benefits and ethical intricacies. While technology stands as a formidable ally in the documentation, analysis, and revitalization of endangered languages, the delicate balance between its advantages and the inadvertent risks of linguistic erosion demands circumspect consideration. The delicate interplay between leveraging technology to safeguard linguistic diversity and the subtle undercurrents of language shift underscores the imperative for a judicious approach. Technological interventions, if wielded without careful calibration, hold the potential to inadvertently contribute to the dilution of linguistic authenticity, as globalized influences permeate traditional linguistic ecosystems. The seduction of major languages facilitated by technology, although providing economic and educational opportunities, poses a concomitant risk of accelerating language decline if not tethered to community-driven preservation initiatives. Ethically, the incorporation of AI in language preservation introduces a complex terrain marked by concerns of cultural appropriation, representation biases, and the exacerbation of existing digital divides. The involvement of external entities in technological developments, coupled with the challenge of ensuring accurate and culturally sensitive representation within AI models, underscores the critical need for ethical frameworks that prioritize community agency, cultural autonomy, and inclusive technological practices. In navigating this multifaceted landscape, a harmonious convergence between technological advancements and community-centric preservation endeavors emerges as imperative. Collaborative initiatives that involve linguists, technology developers, and indigenous communities foster a synergistic relationship, where technology augments rather than supplants grassroots efforts. Such a symbiotic approach not only enhances the efficacy of language preservation but also upholds the ethical imperative of preserving linguistic and cultural diversity.

Ultimately, the trajectory of technology in language preservation is contingent upon the conscientious calibration of its application, recognizing that its potency lies not in autonomy but in complementarity with community aspirations. As we traverse this delicate juncture, it is incumbent upon scholars, linguists, and technology developers to negotiate a nuanced path that safeguards the sanctity of endangered languages, aligning technological interventions with the overarching goal of preserving linguistic heritage for posterity.

References:

1. David Crystal, "Language Death" (2000) - Crystal explores the phenomenon of language death, discussing the reasons behind it and the implications for linguistic diversity.
2. Nicholas Evans, "Dying Words: Endangered Languages and What They Have to Tell Us" (2009) - Evans discusses the richness of linguistic diversity and the importance of preserving endangered languages.
3. K. David Harrison, "When Languages Die: The Extinction of the World's Languages and the Erosion of Human Knowledge" (2007) - Harrison explores the cultural and intellectual loss associated with language extinction.
4. Susan Ervin-Tripp, "Culture, Language, and Thought: Psychological Development in Cultural Context" (1996) - Ervin-Tripp's work delves into the intersection of culture, language, and cognitive development.
5. Lyle Campbell, "Language Documentation and Description" (2011) - Campbell addresses the methodologies and challenges of documenting endangered languages.
6. Lenore A. Grenoble, "Language Policy in the Soviet Union" (2003) - Grenoble explores language policies in the Soviet Union, touching on their impact on minority languages.
7. Leanne Hinton, "Bringing Our Languages Home: Language Revitalization for Families" (2013) - Hinton focuses on language revitalization efforts at the family level.
8. Peter Austin, "One Thousand Languages: Living, Endangered, and Lost" (2008) - Austin provides an overview of the world's languages, including discussions on endangered languages.

-
9. Tove Skutnabb-Kangas "Linguistic Genocide in Education—Or Worldwide Diversity and Human Rights?" (2000) - Skutnabb-Kangas discusses linguistic diversity, education, and human rights.
 10. Salikoko S. Mufwene M. "The Ecology of Language Evolution" (2001) - Mufwene explores the ecological factors influencing language evolution, touching on issues related to language endangerment.