

WILL AI KILL LANGUAGES OR SAVE THEM? THE TWO FACES OF TECHNOLOGY

Siddhi Sunil Bagawade
M.A. English
Chatrapati Shivaji College, Satara
siddhibagawade2002@gmail.com

Abstract

Thousands of the world's languages are disappearing at an alarming rate, even as Artificial Intelligence (AI) grows increasingly capable of understanding, generating, and translating human language. This paper examines whether AI will accelerate language extinction or act as a force for preservation. It argues that AI possesses two contrasting faces. On one side, most AI systems are trained on dominant global languages such as English, strengthening their power while neglecting smaller and endangered languages, which pushes them closer to extinction. On the other side, AI offers powerful tools for documenting disappearing languages, teaching them to new learners, and enabling their presence in digital spaces. The decisive factor is control. When AI is governed by large corporations driven by profit, linguistic diversity suffers. When communities themselves use AI for their own cultural goals, it can become a powerful ally in language survival. The future of linguistic diversity depends on ensuring that AI remains a supportive tool for human communities rather than a replacement for human linguistic traditions.

Keywords: Endangered Languages, Language Death, Artificial Intelligence, Digital Preservation, Language Learning, Community Control, Language Justice

Introduction:

Languages Are Dying, AI Is Rising

Imagine a vast library burning down, one book at a time, with no way to recover the knowledge once it is gone. This is not a metaphorical exaggeration but an accurate description of what is happening to human languages today. Linguists estimate that every two weeks, one of the world's approximately 7,000 languages becomes extinct as its last speakers die. Each lost language represents far more than a collection of words; it embodies unique histories, cultural memories, ecological knowledge, oral traditions, humor, and

worldviews that cannot be fully translated into another language. Language loss is therefore not only a linguistic issue but a cultural and intellectual crisis for humanity as a whole.

At the same time, Artificial Intelligence has emerged as one of the most transformative technologies of the modern era. AI systems now write essays, translate texts, recognize speech, and generate conversations that closely resemble human communication. These systems are rapidly becoming embedded in daily life through smartphones, educational platforms, workplaces, and social media. As AI grows more linguistically sophisticated, it inevitably intersects with the crisis of language extinction. This intersection raises a critical question: will AI accelerate the disappearance of small languages, or can it be harnessed to help save them?

This paper argues that AI is neither inherently destructive nor inherently protective of linguistic diversity. Instead, it is a double-edged technological force whose impact depends on social structures, economic priorities, and ethical choices. Without intervention, AI is more likely to strengthen already dominant languages and marginalize smaller ones. However, when guided by principles of linguistic justice and community ownership, AI has the potential to become one of the most powerful tools ever created for language preservation and revitalization.

The Structural Problem: How AI Marginalizes Small Languages

The first major concern lies in how AI systems are built and trained. Modern AI language models require enormous quantities of data to function effectively. They learn language patterns by analyzing billions or even trillions of words collected primarily from the internet, books, social media platforms, and digital archives. This data is overwhelmingly dominated by a small number of global languages, particularly English. As a result, AI systems become increasingly sophisticated in handling these languages while remaining almost entirely incapable of processing languages with limited digital presence.

This phenomenon has been described as a “rich get richer” effect in digital linguistics. Languages that already dominate online spaces gain even more technological support, while smaller languages are pushed further into invisibility. When a language lacks AI tools such as voice recognition, translation software, spell-checkers, or educational applications, it begins to appear impractical or irrelevant in modern life. Over time, speakers may abandon their native language in favor of a dominant one that offers greater access to education, employment, and digital participation.

Another major factor contributing to language decline is what can be described as the “cool factor” of technology. For younger generations, language is not only a means of communication but also a marker of identity and modernity. When smartphones, social media platforms, video games, and music streaming services operate almost exclusively in dominant languages, those languages come to represent progress and global belonging. In

contrast, heritage languages may be perceived as outdated, rural, or impractical. AI intensifies this imbalance by making dominant languages even more accessible and attractive through predictive text, chatbots, and entertainment platforms, thereby weakening motivation among younger speakers to learn their ancestral tongues.

Cultural Loss Through AI Translation and Representation

Beyond issues of access and prestige, AI also poses subtler threats to linguistic and cultural integrity. Language is deeply embedded in culture, and many words carry meanings that cannot be directly translated into another linguistic system. When AI systems trained primarily on Western languages attempt to translate indigenous or minority languages, they often flatten complex cultural concepts into simplified equivalents. Sacred terms, kinship systems, ecological knowledge, and spiritual beliefs can lose their depth and significance in translation.

For example: a word referring to a sacred tree, animal, or ritual object may be translated into a generic English term, erasing its spiritual and cultural dimensions. Similarly, sentence structures may be altered to fit dominant grammatical patterns, stripping the language of its stylistic identity. While these distortions may appear minor, repeated misrepresentation can gradually undermine the cultural value of a language and encourage speakers to shift toward more “accurate” dominant languages.

Additionally, there is the ethical issue of linguistic data extraction. In some cases, researchers or institutions record endangered languages for academic or technological purposes without ensuring long-term benefits for the communities involved. Language recordings may be stored in inaccessible archives or used to train AI models that communities cannot control or access. This practice transforms cultural heritage into exploitable data rather than a living tradition, reinforcing historical patterns of extraction and inequality.

AI as a Tool for Language Preservation and Revitalization

Despite these risks, AI also offers unprecedented opportunities for language preservation when used responsibly. One of its most valuable roles is as a digital archivist. Many endangered languages exist primarily in oral form, with elders serving as the last fluent speakers. Recording and transcribing their knowledge manually can take decades. AI-based speech recognition tools can dramatically accelerate this process by producing preliminary transcriptions that humans can later refine. This reduces time, cost, and labor while preserving invaluable linguistic material.

AI can also analyze linguistic patterns across large collections of texts or recordings. By identifying recurring grammatical structures, verb forms, and vocabulary usage, AI can assist in the creation of dictionaries, grammar guides, and teaching materials. This is particularly valuable for languages that lack standardized writing systems or formal linguistic documentation.

In the domain of education, AI can function as a patient and adaptable teacher. AI-powered language learning applications can provide personalized instruction, adapting to the learner's pace and weaknesses. These tools can offer pronunciation feedback, interactive dialogues, and culturally relevant examples that make learning engaging and accessible. For children growing up away from fluent speakers, such tools can play a crucial role in reconnecting them with their linguistic heritage.

Making Languages Digitally Alive

AI also helps bridge the gap between traditional languages and modern digital life. Basic machine translation tools, even if imperfect, can enable communities to access healthcare information, legal documents, and educational materials in their own language. Text-to-speech technology can give voice to written texts, benefiting communities with low literacy rates or strong oral traditions.

Importantly, AI can help make endangered languages feel contemporary rather than obsolete. Social media platforms, messaging apps, and digital storytelling tools powered by AI can encourage young speakers to use their heritage languages creatively. When a language is visible and functional in digital spaces, it regains relevance and prestige.

Real-world success stories demonstrate this potential. In New Zealand, Māori-led organizations such as Te Hiku Media have developed speech recognition tools trained exclusively on Māori voices, ensuring cultural accuracy and community ownership. In Hawaii, AI-based grammar checkers and educational platforms support Hawaiian language learners and writers, reinforcing linguistic standards and confidence. These projects illustrate that when communities control the technology, AI becomes a means of empowerment rather than domination.

Ethical Frameworks and Community Control

For AI to genuinely support language preservation, ethical frameworks must prioritize community leadership. Language revitalization efforts should begin with the needs and goals of the speakers themselves, not external researchers or corporations. Communities must

retain ownership over their linguistic data and have the authority to decide how it is used, shared, and developed.

Training community members to manage AI tools is equally important. Sustainable preservation depends not on complex, expensive systems but on simple, well-designed tools that communities can maintain independently. Small-scale applications focused on specific needs—such as story recording, pronunciation practice, or vocabulary learning—are often more effective than ambitious but inaccessible technologies.

Addressing the “cool factor” is also essential. AI should be used to make heritage languages exciting for younger generations through creative digital content such as music, memes, games, and interactive storytelling. When languages become part of everyday digital expression, they are more likely to survive and evolve.

Conclusion: Choosing the Future of Linguistic Diversity

Artificial Intelligence, like any powerful tool, reflects the values of those who wield it. In the hands of profit-driven corporations, AI risks accelerating linguistic homogenization, leaving a world where only a handful of dominant languages thrive. In the hands of communities, educators, and activists, AI can help record endangered languages, teach new generations, and preserve cultural knowledge in innovative ways.

The question of whether AI will kill languages or save them does not have a predetermined answer. It depends on choices made today about ownership, ethics, and inclusion. By prioritizing linguistic justice and community empowerment, humanity can ensure that AI becomes a guardian of linguistic diversity rather than a force of erasure. The survival of the world’s languages ultimately depends not on technology itself, but on the collective will to value and protect every human voice.

References

Crystal, D. (2000). *Language Death*. Cambridge University Press. The classic book explaining why languages disappear and why it’s a crisis.

Kornai, A. (2013). Digital Language Death. *Computational Linguistics*, 39(3), 515–537. Introduces the key idea that a language needs a digital presence to survive in the modern age.

Joshi, P., et al. (2020). The State and Fate of Linguistic Diversity in the NLP World. *ACL Anthology*. Shows data proving AI research overwhelmingly ignores most of the world's languages.

Bird, S. (2020). Decolonising Speech and Language Technology. COLING. Argues that communities must control AI tools for their own languages.

Mozilla Common Voice Project. (Ongoing). An open-source project that lets communities build speech datasets for their languages.

Te Hiku Media (Māori) & ‘Ōiwi TV (Hawaiian). Case studies of Indigenous communities successfully using AI to revitalize their languages.

Bender, E.M., et al. (2021). On the Dangers of Stochastic Parrots. FAccT Conference. Explains how large AI models are built mostly on English data, making them biased and unsuitable for other languages and