A Loss for Words: The Language Extinction Crisis

In a region known as The End of the World, in a small house by the sea, an old woman weaves reed baskets by a smoldering wood fire and listens to the waves crash onto the shore. She looks out the window thoughtfully, hearing the voices of her grandchildren as they play in the cool ocean spray. Her people, the *Yahgan,* navigated these dangerous waters south of the Tierra del Fuego Archipelago in Argentina for over 6,000 years, traveling between the islands in canoes and hunting for seal and shellfish. They numbered almost 10,000 strong when European settlers first arrived 150 years ago (Buenos Aires Times); however, disease, European whalers, and loss of land soon decimated the population, and by 1928, when Cristina Calderon was born, there were very few full-blooded Yaghan remaining, and fewer still who could speak the native tongue (France 24).

As Argentinians began to settle the remote areas of South America once occupied by the indigenous peoples, the Yahgan were forced to abandon their nomadic lifestyle and traditional language. Calderon, however, was raised speaking only Yaghan, learning Spanish from a friend when she was nine years old. She made many efforts in her lifetime to preserve her culture and language—she practiced traditional ceremonies and crafts, worked to create a Yaghan to Spanish dictionary (despite the language originally lacking a written form), even hosting Yaghan language workshops for the locals (Rodriguez). But it was too little too late; when Cristina Calderon died on February 16th, 2022, so too died the last native speaker of the Yaghan language.

But it is not only the Yaghan language that has been lost to time. The United Nations estimates that an indigenous language dies with its last speaker every two weeks. In fact, some calculations assert that up to 95% of the world's 6,700 languages will be extinct or critically endangered by the end of the century (United Nations). This is largely due to worldwide

globalization trends creating an interconnectedness and interdependence between many different cultures and economies through trade, travel, and high-speed information sharing. Previously isolated indigenous groups are now expected to participate in these global affairs, forcing them to learn the dominant language of their region in order to work and produce for the countries they reside in. Unfortunately, globalization does not have the bandwidth for all of the individual cultures that comprise it, so beyond learning the language, these indigenous groups are also expected to assimilate into the more dominant cultures, facing harsh discrimination if they fail to do so.

Therefore, it is not enough for these speakers to simply become bilingual and retain their own culture; they must, for the convenience of a larger society, abandon their traditional languages and cultures entirely in order reap the rewards from society and avoid the consequences it would otherwise impose. With this loss of traditional languages, however, comes the inevitable and devastating loss of history and culture that can only be properly communicated through the language in which it occurred, leaving millions of stories untold, and millions more never to be written.

But while some languages have been lost to time, researchers are determined to slow—or perhaps reverse—the decay of the ones we have left. One way they have begun to do this is through one of the most powerful tools of the 21st century: artificial intelligence. Large Language Models (LLMs) are a form of AI in which a computer program is fed large amounts of human text until it is able to recognize, interpret, and eventually produce human language. Its most common use is for generative AI—users input a prompt, and the program predicts a probable response (Cloudfare). But at Dartmouth University, computer science students are beginning to realize LLM's potential for language preservation. One of these students, Ivory Yang, recalled learning a few words of Nüshu from her grandmother when she was a girl. Nüshu was a form of

writing used 400 years ago by Yao women in the southern Chinese province of Hunan to communicate in secret (Lofthouse). Many of these writings have been lost or destroyed, but those that have been salvaged were scanned by researchers and paired with their corresponding Chinese translations. Using a 500 piece dataset and only 35 sample phrases , the LLM was able to accurately translate novel phrases over 21 years after the last known native speaker of the language passed, representing a monumental victory for language preservationists (Barath). Similar language models have since been used to revitalize other languages, such as Wôpanâak in the United States (Lamentillo), BriBri and Cabécar in Costa Rica, and Māori in New Zealand (Sanusi).

Unfortunately, many languages on the verge of extinction are spoken in areas where internet access is relatively limited. So, in cases where AI technologies are impractical or unavailable, some governments and teacher associations in under-resourced areas have also begun launching endangered language initiatives in schools. There have been efforts made in Oklahoma (Cherokee), Hawaii (Ōlelo Hawai'i), and even the small town at the End of the World, Ushuaia (Yaghan), to begin teaching native languages in public schools, hoping to revitalize the next generation of speakers and encourage them to maintain their cultural identity (Flannery).

However, both AI and educational initiatives have their limitations. The prevalence of the internet has introduced a new category of languages now considered *digitally disadvantaged*. This term means that a language is not supported on most digital platforms—and it is a term that can be applied to over 98% of the world's languages (Zaugg). This spells disaster for language preservation in an increasingly digital world—so much of our world's commerce, information sharing, and social relationships occur online, and a lack of support for smaller language communities has either kept them off the internet (unlikely) or caused them to assimilate into the larger language groups. In fact, a 2008 study by UNESCO found that 98% of internet webpages

were written in just 12 major languages (Diki-Kidiri). This limits the practicality of AI as a language preservation solution because the LLMs used for language preservation must be trained using large amounts of language data that simply does not exist online for these digitally disadvantaged languages. Even the AI models designed for multilanguage use frequently mistranslate less popular languages and display an English-first bias, even to the point of producing illogical answers, fabricated words, and sometimes total gibberish (Deck). Many languages, like Yaghan, do not have a written form, and therefore have so little data available even outside of the digital world to ever teach computer programs the language, much less other people. That is truly the caveat of AI as a language preservation method--AI can potentially record an endangered language, maybe even learn to produce it, but it cannot save it, for the simple reason that it cannot speak it.

Language can only be possessed and protected by speakers. By humans. The language extinction crisis is not a result of endangered languages not having functional AI LLMs, it is a result of the millions of children that are not learning the language of their grandparents. Historically, public education offered to Indigenous groups was not offered in their traditional tongue, but rather in the language of the government providing the education under the impression that assimilation into white culture was the best (albeit cheapest and easiest) outcome for Indigenous youth (Toth). But in surveys across First Nations groups in Canada, it was revealed that education on their culture, language, and history was an area of major priority for many students (Canada), so it is not a lack of interest that has fostered a lack of learning. Rather, it is a lack of focused initiatives, adequate resources, and qualified staff (Flannery) that would require large scale changes in reservation education to satisfy--but these are the changes that must take place. The *only* way to protect language diversity on our planet is to teach the next generation to speak the language of their ancestors. It is not convenient, but focused efforts by

parents and elders to prioritize the education of traditional languages for their children is what keeps a culture alive and well.

Endangered languages must be taught. Without intervention, there is little hope for the future of language diversity, and with something so rich with knowledge, so intensely personal as human language, no effort can be spared. Software and website developers must work to create space online for all languages, alphabets, and syntax. Linguists and researchers must continue to harness AI for all that it's worth, before it's too late. Governments must provide resources for interested language learners in public education, even in schools outside of the smaller speech communities. Parents must teach their children to speak in the voice of their forefathers. Even in my own field of speech pathology, efforts should be made to provide services to those in threatened speech communities in order to encourage and support the next generation of speakers in every way possible, because language is so much more than words. It's woven so intricately into the fabric of humanity that it is impossible to separate the two. Language is the medium for our thoughts, for our conversations, for our relationships. It's our window to the past and our prayers for the future. For an old woman at The End of the World, it was a precious reminder of the people that came before her, a life's work, and a final legacy. And for the sake of the 88 million people around the world whose languages may disappear in the weeks to come (Lu), I hope it is something important enough to protect.

Works Cited

Barath, Harini. “Language Preservation Efforts Get an AI Boost | Dartmouth.” *Dartmouth.edu*, 9 Apr. 2025, home.dartmouth.edu/news/2025/04/language-preservations-efforts-get-ai-boost.

Canada, Government of Canada; Indigenous Services. “Let’s Talk On-Reserve Education: Survey Report.” *Www.sac-Isc.gc.ca*, 26 Oct. 2017, www.sac-isc.gc.ca/eng/1509019844067/1531399883352?ref=hir.harvard.edu.

Cloudflare. “What Is a Large Language Model (LLM)?” *Cloudflare.com*, 2022, www.cloudflare.com/learning/ai/what-is-large-language-model/.

Deck, Andrew. “We Tested ChatGPT in Bengali, Kurdish, and Tamil. It Failed.” *Rest of World*, 6 Sept. 2023, restofworld.org/2023/chatgpt-problems-global-language-testing/.

Diki-Kidiri, Marcel. “Securing a Place for a Language in Cyberspace.” *Unesco.org*, 2025, unesdoc.unesco.org/ark:/48223/pf0000149786_eng. Accessed 24 Apr. 2025.

Flannery, Mary Ellen. “Educators Work to Preserve Native Languages | NEA.” *Nea.org*, 2023, www.nea.org/nea-today/all-news-articles/educators-work-preserve-native-languages#91876. Accessed 24 Apr. 2025.

https://www.facebook.com/FRANCE24.English. “Chile’s Last Yaghan Speaker Dies Aged 93.” *France 24*, 16 Feb. 2022, www.france24.com/en/live-news/20220216-chile-s-last-yaghan-speaker-dies-aged-93. Accessed 24 Apr. 2025.

Lamentillo, Anna Mae Yu. “AI: The Unexpected Hero in the Battle to Save Dying Languages.” *Apolitical*, 2024, apolitical.co/solution-articles/en/ai-the-unexpected-hero-in-the-battle-to-save-dying-languages-844.

"Last Native Speaker of Chile's Indigenous Yaghan Language Dies Aged 93 | Buenos Aires Times." *Www.batimes.com.ar*, 16 Feb. 2022, www.batimes.com.ar/news/latin-america/last-native-speaker-of-chiles-indigenous-yaghan-language-dies-aged-93.phtml.

Lofthouse, Andrew. "Nüshu: China's Secret Female-Only Language." *Www.bbc.com*, 1 Oct. 2020, www.bbc.com/travel/article/20200930-nshu-chinas-secret-female-only-language.

Lu, Marcus. "All the World's Endangered Languages, by Country." *Visual Capitalist*, 15 Mar. 2024, www.visualcapitalist.com/cp/all-the-worlds-endangered-languages-by-country/.

Rodriguez, Aliciamarie. "World's Least Common Language, Yagán, Dies out Completely." *Guinness World Records*, 23 Mar. 2022, www.guinnessworldrecords.com/news/2022/3/worlds-least-common-language-yagan-dies-out-completely-696065. Accessed 24 Apr. 2025.

Sanusi, Tife. "Reviving Lost Tongues: How AI Battles Language Extinction." *Deepgram*, 17 Aug. 2023, deepgram.com/learn/language-ai-cultural-preservation.

Toth, Katalina. "The Death and Revival of Indigenous Languages." *Harvard International Review*, 19 Jan. 2022, hir.harvard.edu/the-death-and-revival-of-indigenous-languages/.

United Nations. *Indigenous Languages Language Rights of Indigenous Peoples*. 2018.

Zaugg, Isabelle A., et al. "Digitally-Disadvantaged Languages." *Internet Policy Review*, vol. 11, no. 2, 11 Apr. 2022, policyreview.info/glossary/digitally-disadvantaged-langu