



Harnessing AI to Build the Ecology of Indigenous Thoughts: An Interview with Michael Running Wolf by Isabelle Alice Zaugg on 4 April 2025

MICHAEL RUNNING WOLF

ISABELLE ALICE ZAUGG 

**Author affiliations can be found in the back matter of this article*

**SPECIAL COLLECTION:
THE DIGITAL LIVES
OF GREATER SOUTH
LANGUAGES**

**ARTICLES – DIGITAL
MODERN LANGUAGES**



ABSTRACT

Both Indigenous researchers and Indigenous languages are under-represented in the emerging field of Artificial Intelligence (AI). Michael Running Wolf, founder of Indigenous in AI, the Lakota AI Code Camp, and the First Languages AI Reality project (FLAIR), is working to change that. In this interview with communication scholar Dr Isabelle Alice Zaugg, he shares his experiences building expectations of accountability among AI researchers working on Indigenous languages and what motivated him to train a new generation of Indigenous coders in rural South Dakota. He discusses his work to develop AI tools that support language learning in Indigenous communities across the full spectrum of North American languages that have co-evolved together. He also proposes nuanced, collective strategies to balance data sovereignty with digital access to Indigenous languages. Motivated by his desire to “build and rebuild the ecology of Indigenous thoughts”, Michael Running Wolf reflects on his multifaceted approach to changing the narrative for endangered Indigenous languages through AI that “creates speakers, not technologies”.

CORRESPONDING AUTHOR:

Isabelle Alice Zaugg

Independent Scholar

isabellezauggphd@gmail.com

TO CITE THIS ARTICLE:

Wolf, Michael Running and Isabelle Alice Zaugg 2026 Harnessing AI to Build the Ecology of Indigenous Thoughts: An Interview with Michael Running Wolf by Isabelle Alice Zaugg on 4 April 2025. *Modern Languages Open*, 2026(1): 4 pp. 1–7. DOI: <https://doi.org/10.3828/mlo.v0i0.578>

Isabelle Alice Zaugg: You have devoted much of your professional life to fostering the reclamation and vitality of Indigenous languages through AI. Two aims you mentioned in your 2023 TEDx talk “Why First Languages AI Can Be a Reality” that stood out for me are 1) “We need to change the narrative for endangered languages” and 2) “We can build and rebuild the ecology of Indigenous thoughts through AI.” Could you please expand on what those intentions mean to you?

Michael Running Wolf: Really good questions. When I say narrative for languages, it’s always the stark story of, “Woe is us, the languages are expiring and there’s nothing we can do other than document.” It’s sort of a continuation of the romanticization of the “noble savage” that was very common in anthropology and broader research until recently, which still exists to some extent. Anthropologists were not motivated by the good of the community, but rather took the approach of “We need to preserve this knowledge because it’s about to go extinct” – like the dodo or the carrier pigeon – versus, “What can we do to protect and ensure that the languages thrive?” It was basically a collective decision by anthropological science. More recently, there have been reforms to make sure anthropological research is not exploitative and rather is co-creative, like *collaborative participatory research*. But this approach is still pretty new and not as widespread as it should be.

What is the objective of AI? How does AI fit into all of this? When I say ecology, I want a broad solution that doesn’t just serve my language. I want my mother’s language and father’s language to thrive into the future, but it’s not sufficient if I only work for the protection of my tribe’s language. That would be a minor victory, because it’s the collective interaction between different tribes and different languages that brings the full wealth. These languages co-evolved, such as the Arapahoe and the Cheyenne. The same with the Cheyenne and the Lakota. These languages shared culture with each other and that has happened at a macro scale. But this is not evident in the literature.

The literature tends to look at languages in isolation, it doesn’t capture that these languages are interacting in common spaces, through families and in shared spaces such as powwows. We can’t only save Lakota, or Arapahoe, or Diné. Therefore, we can’t build AI using a linguistically isolated approach. I’m not interested in building the Lakota AI project or the Diné AI project. I want a solution that contributes to the overall health of this linguistic system rather than working piecemeal.

Also, *saving* is the wrong paradigm. How do we bring these languages along? If a new technology does not support multilingualism, languages suffer. Like the DARPA program, which first was an exclusively American thing. Technology, when it’s introduced, there’s a lot of unconscious, unintentional biases. When the Internet was created, it was an American invention, it was American infrastructure to the exclusion of other nations, and you had to speak English to use it. It wasn’t done on purpose, yet until today the lingua franca of the Internet is English.

What if the Internet was invented by Serbians or Italians? Or what if it was invented with an eye to multilingualism? AI is starting to become hypey and it’s going to transform our business economies. And if it only speaks European languages, it will be another layer of exclusion. The Internet doesn’t really support Lakota. What if from the get-go the language of AI wasn’t exclusionary to only a few languages?

IAZ: Can you share your motivations behind founding Indigenous in AI, an initiative that focuses on speech-recognition tools for the Indigenous languages of North America?

MRW: The objective was, I started building this research, I built the team, did fundraising. It’s very clear there’s not enough of us doing this work.

I happened to be invited to the largest AI conference in the world in 2019. I had a friend of a friend invite me. When I was there, I didn’t yet know what it was at the time, it was very early stages. The point was I went there with the eye of finding other native AI researchers. I went to all the poster sessions and all the talks where maybe there would be native applications. NLP (natural language processing) technologies for a lot of low-resource languages were being demonstrated, but none of them were Indigenous. Georgia Tech, MIT, top ten universities were working on low-resource languages, but the research was being conducted without the awareness of the communities involved.

There was some work on Cherokee being conducted, and I was bugging them, “Does the Cherokee tribe know about this?” When you do *research* on a tribe, you’re expected to complete a research ethics review, but they weren’t even doing that. They said “I found it on the Internet.” I told them, “I still think it’s important that you go and ask them: Do you know your data is being included in every AI research?” I was asking very obvious questions about, “Where is the tribal representation?” The research we had stopped the academics from doing, the corporates and AI developers were now doing. ML (machine learning) researchers in the very same institutions are not being held to the barest minimum of ethics requirements simply because it’s a new space. They’re not used to this [expectation].

But the main point for me was, “Where are the rest of us?” I happened to be raising so much of a ruckus that one of the conference organizers introduced himself. People got to know me very rapidly, and I knew what I was talking about because I was in the space. The person running the conference said, “You should start an affinity group for natives.” They had Black in AI, Women in AI, etc. The objective was to find other people doing research who were Indigenous, and number two, to bring the message to the scientists on the forefront of these technologies that there were ethical concerns being caused by these types of research. It’s slow – I don’t have all the time in the world – but among my peers, it’s now normal that corporations at least talk to communities. They might not stop what they are doing, but they’ll try. It’s a minor success.

But I attended the conference primarily to find others, and there are very few [Indigenous AI developers]. I quickly realized we don’t exist. Few and far between; once in a while I find a new one. There’s about a dozen today. I know ten and I’m only hopeful there’s two more I don’t know. There’s only ten across the USA, Mexico, and Canada. That kind of inspired me to the next phase of this work, which was thinking we need to create our own pipeline. There’s a gap in the pathway, where Indigenous youth do not think tech fields are a career path.

I knew this from my work at Amazon. One of the things we focused on was recruitment, getting Native Americans into Amazon. We only found five people across the entire country, three from MIT, one from Stanford, the usual universities. And so, I knew before I walked into NeurIPS (Conference and Workshop on Neural Information Processing Systems) that there weren’t many, but I came to find there were none.

The next part of the journey was organizing a camp with help from friends from academia and the organizers of NeurIPS – a lot of people helped. I concluded among friends, “We need to train our own kids. We’re just going to do it; no one else is going to do it.” We had to create a model. We take twenty kids from South Dakota, stay local, find a college partner, teach them basic coding and AI frameworks. We give them laptops, feed them three meals a day, and they can do it. We were told by education theorists, academics, that it was going to be impossible, that the kids’ brains would fall out the side of their head. Kind of like the idea that women should not read because their brain would get too hot [*laughs*]. We just did it.

We put together a curriculum and tried to balance being too technical and being interesting. Six of us decided we’ll do this with high standards. And it worked. None of the kids had prior software experience. These were rural, poor kids from South Dakota. One kid wasn’t eating every day and was very excited we were feeding them. Poor, rural Lakota kids. None of them had owned a laptop. One pushback from our funder was, “Why aren’t you going to a lab?” There was no lab in South Dakota that could host us. We didn’t want to fly them to Stanford or MIT. I wanted to show the kids that we could do it here, with your bare feet touching your own land.

I’m a product of the brain drain, by the way. My mother worked for HP (Hewlett Packard) in Aurora, Colorado, driving a new car every year. She decided she didn’t want that for us and went back to Birney, Montana, and raised us here. That’s the model I wanted to share with the kids. All these logistics came slamming together, and it worked. We were an amazing team, everyone gave 110%. We stayed up until midnight every day. We were watching the kids, seeing how they responded to the curriculum. There was no model of how to teach rural kids AI in the sticks with no Internet. Our goal wasn’t for them to pursue a career as computer engineers, rather that they go to college and when they take their first computer course, they aren’t as scared. They’ve done it before.

When I went to university, my first semester there were 200 computer students, and by the next semester it was half, and in the end only ten graduated. Plus, I was brown and no one looked at me. Same with women. Every single kid who graduated from the Lakota AI Code Camp is now in a four-year institution pursuing computer science related fields, maths, etc. I think that's amazing. These kids just needed to be shown they could do it. There's going to be a wave of AI researchers soon if we keep it going. Indigenous in AI started off with a realization that we don't exist. It was kind of created out of the realization that we needed to create our own people. We have a website you can check out: <https://Lakota.AIcode.camp>.

IAZ: Through the First Languages AI Reality project, or FLAIR, you are working on a speech-recognition model for the 200 endangered Indigenous languages in North America. You mention your aim is to "create speakers, not technologies". On the Indigenous AI website you state, "Imagine Virtual Reality worlds populated by intelligent Hawaiian language speakers wanting nothing more than to teach you a new language. Everyone needs an infinitely patient Indigenous personal language teacher." Tell me how you see Indigenous AI changing the game in terms of Indigenous language acquisition.

MRW: I think that was very *pie-in-the sky*, obviously. That's not going to happen any time soon; there are a lot of technical barriers. Our near-time objective is very modest. We are trying to create an AI that is basically a grammar and phonetic recognizer and can judge your phonetic and grammar mastery of the language. These languages are complex morphologically and it's very hard to get good practice when your language has very few speakers. Every tribe has advocates and teachers who are trying to keep the language thriving, and there are plenty of enthusiastic students who want to learn. But the problem is you can't practise at home. So it becomes the language of *ceremony*. Two-thirds of us live off reservations, so you don't have access to speakers. It's not a lack of motivation. Languages are not losing speakers because of lack of motivation, but it's a logistical difficulty. That's where AI is coming in. What if you can provide a tutor, not even an intelligent tutor? It's a difficult position for a second language speaker. You're coming in with a lifetime of speaking English or a certain language, which is not compatible with your home language.

That's a big emotional barrier to overcome, because all humans are shy about learning a new language, except children. Native people have the added barrier of not feeling authentic. So, the vast majority of Indigenous want to learn, but we're not getting there. So we talked to friends and family and experts, went to conferences. The common theme was just not enough speakers. You go to class, get taught, you go home, the class ends, and it kind of fades. On top of that, you need to have the confidence to continue the practice, but sometimes you can't continue this because of emotional anxiety and cultural trauma.

AI can be an intervention, as you're in the privacy of your own home, in the privacy of your own space, it's privacy-centred, no recordings. AI is dumb, it doesn't care how you say things. It's an inference tool – it's not *judging* your skills, just providing arbitrary metrics. We can also tune it to what the community wants it to do. At first it can be loose, and later we can tune it tighter. We can use bias in a mathematical way so the student feels comfortable. We can tune it for gender. At the end of the class, they can speak to elders in a way that the community deems sufficient to what the community wants, and hopefully motivate them to continue learning. You can't become fluent in one class.

Maybe in five years AI could be your tutor, that could be the last resort. Far in the future, if people stop speaking, if we see communities with cheap computers, cheap Raspberry Pis, and in the future if people want to start speaking the language again, there's a little cedar box that can get you started. But that's far in the future. We're trying to focus on the practical needs we have today.

IAZ: How are you developing speech recognition when working with highly polysynthetic languages, such as languages in the Wakashan language family?

MRW: It's an ML problem. We're doing novel science, unfortunately. Mostly because it's expensive and long, no one's done it. Other people are working on the same problems, so we're not alone. We're standing on the work of CRIM (Computer Research Institute of Montreal) and several others. We're standing on computational linguistics research to do this. We're also borrowing the new strategies of *generative AI*, certain strategies that went into generative AI

that help to advance the science. We're not building a UFO, we're standing on the shoulders of giants, as usual.

IAZ: You are an advocate for data sovereignty within Indigenous communities. Why is this important?

MRW: I think we need to consider treaties and realize that our data is like our land, it's our heritage, it's our right to have true sovereignty – not just permission but *agency*, control at a fundamental level. It's like our heritage or our land. You don't just give away land. You sign a contract, there's a lease. In intellectual property (IP), it's the same. Instead of land, it's data, it's ideas. There's an entire economy based on IP and patents for Disney movies, art. We Indigenous need to treat our data the same way. If a researcher wants to do research on your land, digging holes and starting to do DNA analysis, of course there's going to be a contract. They can't come in with a bulldozer and tear down Grandma's azalea tree. In the AI age, our IP is valuable, and we need to be compensated for the research being conducted on our data. Peter Lucas Jones, who's Māori, he's a big advocate of treating data like our land. Researchers who are taking our data are basically replicating land theft, fundamentally.

IAZ: Which unfortunately we see widespread in the AI field in general.

MRW: Yes, our texts, our posts are all being used to train models. We've sort of normalized that for Indigenous data. Just because it's widescale doesn't mean it's okay.

IAZ: How do you balance the value you put on data sovereignty with your aim of pushing Indigenous languages and cultures into wider and extended spaces such as the metaverse?

MRW: It's a bit of a contradiction, isn't it? The metaverse – which was not invented by the company Meta – it dates back to the 1980s. When I was a kid I was a sci-fi nerd, and there was a book called *Snow Crash*. The author is Neal Stephenson. My dad was into sci-fi. We are in the middle of nowhere, we had nothing to do in the winter except look out the window and read. He would buy very pulpy sci-fi books from second-hand bookshops. You touch the words and they'd smear off. The idea of "plugging in", it goes back to what I was talking about with the Internet. It's not here yet, but the era of AI is coming. The era of VR is coming. And I can see the edges of it appearing, and obviously there's a lot of logistical barriers, like putting on the headsets. They are uncomfortable, they're hot, a lot of people puke.

The Internet is one big privacy violation, but we need to find a way to exist. GDPR (General Data Protection Regulation) needs to be a global phenomenon, the EU (European Union) privacy law. They ignored that. We already exist on Facebook, we're already pouring a lot of our cultural data on to Instagram, so the risk is always going to be there. But it's also how our communities connect, how we plan. Seventy per cent of people connect to those who live on the rez [reservation] through Facebook. I wish these systems had humanity at their forefront, but they are also free and open and commercially viable. You have to balance both. Obviously there has to be decorum; don't put your sacred data on there. Reserve some things to be offline.

IAZ: So is striking that balance a matter of digital literacy?

MRW: I think we need that now. I have elders posting their mobile phone numbers on public forums, trying to get my attention. Elders don't really realize the risks, especially in the age of AI. Every blog post is contributing to generative AI. Unfortunately, the cat is out of the bag. Let's apply the lessons learned to at least avoid the worst expectations in the new technology.

IAZ: In the Hopi community, there have been long-standing rules prohibiting outsiders from learning the Hopi language. For the Navajo code talkers, their language *illegibility* to outsiders was an asset in warfare. Other Indigenous communities may be concerned about digital surveillance of their communications. Is linguistic inclusion in digital spaces a universal good, or are you considering the need for Indigenous community decision making that determines which languages should be included, and which decline to be included?

MRW: I think it needs to be the right of every community to decide to what extent their language is exposed to outsiders. I understand the Hopi perspective. The Cheyenne, my mother's tribe, has a similar perspective, which is very different from my father's tribe the Lakota. It has policy implications, such as who has the right to conduct research and whether they can post their data for anyone to use and look at. The Inuktitut dataset – I'm sure you know about it – there's

a very large dataset for AI [in the Inuktitut language], but we're not touching it. It would be easy, I have the links, but I'd only do it if I had the permission of the community.

There's a lot of public datasets out there, such as the Hawaiian newspaper dataset. It's best to ask permission [from the community] even if it's data that's already publicly available. But it doesn't stop communities that want to share their data. A lot of the data for the FLAIR project will be open sourced. The goal would be to support [each community's goal]. What if the Hopi wanted to create their own AI using our technology but wanted to keep it private? We wouldn't even be involved. We would show them how to use the technology, and shake their hands and say "Good luck." Whereas for the Lakota kids, I think there will be open-source AI everywhere. That's up to the decision of the community, not me.

The Inuktitut dataset was created by the government up north near the Arctic. NLP researchers use it ... but we are not touching it.

With FLAIR we won't open source the data, but we'll open source the tools. That way people can use the technology as they see fit, contributing to the larger solution.

IAZ: Is there a trade-off between linguistic inclusion and accessibility versus data sovereignty? There are obvious benefits to open-source and free use datasets, for example, but are there any risks there as well? I am interested to hear your thoughts on what is to be gained and what potentially might be lost through the inclusion of Indigenous languages in the digital sphere.

MRW: I think it's got to be more nuanced. Currently the paradigm is, it's open source everywhere or it's secret and offline. We need infrastructure for a more nuanced perspective, because if your language is completely offline and not accessible, your community suffers. If your language is completely accessible, it's easy to exploit. Right now, you only have two choices. There's little nuance. I can give you a flash drive or an intranet. An international third path needs to be created where you can give access to stakeholders while also protecting the data. What if your data was accessible securely through trusted partners through a trusted verification system, where you can have access to a dataset? Someone must be doing it. Some places in Canada you have to be a member of a community to access the data store. To get access you have to make a request to their IT administrator. What if that approach was made available through a broadly accessible platform?

As FLAIR matures, I see the necessity for a data co-op to be attached to FLAIR. We are going to be entrusted with the safety of a lot of people's data. I see the need for the creation of a new entity for the data, a guardian to protect it. It's too easy to lose the data. You have to be conscious about good data hygiene to protect data from being lost. I don't think that's the domain of FLAIR, I think it's some NGO governed by thought leaders and elders in the space, and FLAIR would be a client. Or if you want to work with FLAIR, first you have to upload your data to a data co-op and then FLAIR can access it.

This would also protect the data from a legal perspective. It accomplishes two tasks, if you want to share with corporations, but you also want to protect the data through legal means. It's a bit controversial, because why involve Western legal frameworks? The general idea is, "What if there was an institution where you upload your data and outline your desires?" We can even set a price for someone to touch the data. You set the rules and the institution would be in charge of administration of the data for free, plus overheads. That doesn't really exist. Or maybe you want to entrust your data to a data trust and that's all they do – keep it offline and safe.

IAZ: Thank you so much, Michael Running Wolf. Is there anything you wish I had asked? What last thought would you like to leave readers with?

MRW: Your questions were pretty thorough. My favourite colour is blue. I like long walks on the beach with my wife. I'm a Scorpio. I can't think of any others!

AUTHOR AFFILIATIONS

Michael Running Wolf

Isabelle Alice Zaugg  orcid.org/0000-0002-6079-7897

Independent Scholar

Running Wolf, Michael. "Why First Languages AI Can Be a Reality." TEDx talk, 28 June 2023. <https://www.youtube.com/watch?v=Omp3X-FXdLs>. Accessed 27 October 2025.

TO CITE THIS ARTICLE:

Wolf, Michael Running and
 Isabelle Alice Zaugg 2026
 Harnessing AI to Build the
 Ecology of Indigenous
 Thoughts: An Interview with
 Michael Running Wolf by
 Isabelle Alice Zaugg on 4 April
 2025. *Modern Languages
 Open*, 2026(1): 4 pp. 1–7. DOI:
<https://doi.org/10.3828/mlo.v0i0.578>

Published: 11 March 2026

COPYRIGHT:

© 2026 The Author(s). This
 is an open-access article
 distributed under the terms
 of the Creative Commons
 Attribution-NonCommercial-
 NoDerivatives 4.0 International
 License (CC-BY-NC-ND 4.0),
 which permits unrestricted
 distribution, and reproduction
 in any medium, provided the
 original author and source are
 credited, the material is not
 used for commercial purposes
 and is not altered in any way.
 See <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Modern Languages Open is a
 peer-reviewed open access
 journal published by Liverpool
 University Press.