Prospects and pitfalls in using Artificial Intelligence (AI) for Algonquian revitalization

Caroline Running Wolf
(FLAIR; Crow), caroline@firstlanguages.ai
Michael Running Wolf
(FLAIR; Northern Cheyenne / Lakota / Blackfeet), michael@firstlanguages.ai

First Languages AI Reality (FLAIR) is an Indigenous-led initiative founded by Michael Running Wolf (Northern Cheyenne / Lakota / Blackfeet), Caroline Running Wolf (Crow), Dr. Shawn Tsosie (Navajo / Little Shell), and Dr. Conor Quinn. The team combines decades of experience in digital language technology development, effective community-centered language teaching and learning and project management as well as Indigenous data sovereignty.

This talk will outline major aspects of Artificial Intelligence (AI), address systemic barriers to Indigenous-led AI, and attempt to reach beyond the hype of AI to discuss its practical applications for community-led language revitalization.

Indigenous languages and the traditional knowledge encoded within are a valuable resource, especially to academia and big tech mining the internet to train generative AI. FLAIR is working to disrupt that exploitation by modeling ethical AI research centered around the needs of the partnering Indigenous community, respecting Indigenous ways of knowing, and specifically creating protocols and agreements to ensure that intellectual property rights, and rights to the data itself, stay within the community in accordance with Indigenous data sovereignty—alongside building community-internal capacity to collect this kind of data and develop relevant tools themselves.

Conventional approaches to AI training use extremely large sets of data for major languages. No such datasets exist for Algonquian and other Native American languages. FLAIR has developed a multifaceted approach which drastically decreases the data requirements for Automatic Speech Recognition (ASR). FLAIR's immediate focus is to identify solutions for a specific set of Indigenous languages in North America—having started with Wakashan languages (Makah, Kwak'wala), we are now looking to leverage decades of team-member work with Algonquian-language reclamation efforts to offer this approach to Algonquian-speaking communities. FLAIR's approach will make ASR accessible and attainable for thousands of under-resourced speech communities worldwide.

Voice AI for Indigenous languages can power interactive tools that reinforce in-person language learning yet these technologies also pose risk to sensitive community datasets. This talk provides the speakers' perspective on and pathways to responsible AI development for Indigenous communities.