

# Project Description

## Introduction

There's growing interest in using AI tools to support endangered language documentation and revitalization. While much of that interest is well-meaning and some of the tools show real promise, the field has not yet done the work to think through what responsible AI engagement with Indigenous languages should look like. AI is now entering a space long shaped by histories of linguistic extraction. Without thoughtful coordination, current efforts may unintentionally repeat earlier challenges in language documentation, such as using unvetted data, developing tools without community consultation, or emphasizing academic outputs over community applicability. We propose a workshop to convene language teachers, documentation experts, and AI researchers. The goal is to surface shared concerns, highlight best practices, and collaboratively define what responsible, collaborative AI engagement should look like in revitalization settings. This is a formative moment for AI and Indigenous Language Work. The field of language documentation has undergone a significant ethical transformation over the last several decades. What began as a primarily academic and extractive discipline has gradually moved toward more collaborative, community-led models. Today, linguists and archivists increasingly work in partnership with Indigenous communities, co-authoring materials, co-owning data and aligning research priorities with the goals articulated by language experts and educators in the community. Major funders (including NSF's DLI/DEL programs) have helped lead this shift through explicit attention to ethics, community involvement, and long-term sustainability. AI, by contrast, is just beginning to enter this space, and it is doing so rapidly, often without the benefit of that hard-earned wisdom. Many current projects involving AI and Indigenous languages operate with minimal community input, little clarity around consent or data ownership, and few long-term commitments to reciprocity or accountability. While some initiatives are driven by community-based organizations and local experts, these remain the exception rather than the rule. This disparity in maturity presents both a risk and an opportunity. Without action, narrow or externally driven approaches to AI development could quickly become normalized. But if we act now, we can leverage existing models from modern language documentation and revitalization efforts to shape a more just and sustainable future for this work that opens new avenues for rigorous, innovative AI research that would be impossible without grounded, collaborative relationships.

## The Promise Of AI

Despite the real risks and ongoing challenges, there is compelling evidence that AI can play a constructive role in supporting endangered language documentation and revitalization if developed and deployed with care. A growing number of projects have demonstrated that AI tools, including large language models (LLMs), can be adapted to work in extremely low-resource settings when grounded in community knowledge and linguistic expertise. For example, LLM-assisted rule-based translation systems have enabled new revitalization-oriented tools for languages like Owens Valley Paiute, even in the absence of publicly available corpora [1]. Morphological segmentation tools that combine unsupervised and LLM-based methods have been shown to produce more linguistically informed outputs than standard models alone [2]. Other studies have explored how LLMs can be adapted to previously unseen languages using small grammars [3] or limited parallel data [4], offering new paths for supporting low-resource languages without relying on large datasets. Automatic speech recognition (ASR) for under-resourced and polysynthetic languages has likewise seen progress, though challenges remain. Some recent work shows that no single ASR architecture consistently outperforms others across endangered languages [5], while others have found that modern

ASR systems struggle with the morphological complexity typical of many Indigenous languages [6]. These findings underscore both the technical challenges and the scientifically rich questions they raise.

Equally important, there is growing consensus around the idea that community-defined goals must shape AI development. Liu et al. [7] emphasize that technological innovation must not overshadow the needs, values, and lived experiences of the communities these systems aim to support. These examples represent just a sample of the innovation emerging at the intersection of AI and endangered language work. They show that when AI is aligned with community priorities, it can help address longstanding challenges in documentation and revitalization. Our proposed workshop seeks to build on this momentum by bringing together the expertise needed to ensure that future AI tools are technically effective, well-validated, and developed in close collaboration with language stakeholders.

## **Identified Gaps And Challenges**

While interest in applying AI to Indigenous language revitalization is growing, the field currently faces several challenges that must be addressed in order to ensure responsible and community-centered progress.

### **Lack Of Cross-Disciplinary Dialogue**

Despite growing interest in “Indigenous AI,” there are few formal spaces where community members (such as learners and teachers), documentation experts, and AI researchers come together to share insights and concerns. Existing gatherings in the Indigenous AI space like “Indigenous in AI/ML” at NeurIPS [8], AmericasNLP [9], and ComputEL [10] tend to center technical researchers (including some who are Indigenous and many who work with Indigenous communities), but do not explicitly include participants with lived experience in community language work. Conversely, the language documentation field (especially the DLI/DEL community) has long emphasized collaborative, community-led approaches to research. Scholars have articulated concrete models for partnership-based documentation that prioritize reciprocity, co-authorship, and community-defined goals [11], and frameworks like the CARE Principles for Indigenous Data Governance [12] offer specific guidance for ethical data use grounded in Indigenous priorities. These traditions have been further formalized through initiatives like the DLI Community of Science [13], which brings together linguists, technologists, and Indigenous researchers to “promote sustainable and impactful cross-disciplinary language documentation research”. Notably, however, AI researchers have been almost entirely absent from these conversations. Despite this robust foundation, many in the AI community remain unaware of existing models for ethical, community-centered language documentation and revitalization work. Bridging this gap is one of the central goals of our proposed workshop: to connect emerging technical work with decades of experience in language documentation and collaborative practice.

### **Emerging Harms In AI Practice**

We are already seeing signs of AI systems being developed for Indigenous languages without consent or meaningful consultation, without fluency validation or linguistic grounding, and with limited regard for the social and pedagogical consequences of deploying flawed tools. These harms are not theoretical. In low-resource language settings, poorly designed systems can have long-term ripple effects, introducing errors into classrooms, shaping pedagogical practices, and undermining community trust. Because fluent speakers are often few in number, even small inaccuracies can spread quickly, especially when baked into tools seen as authoritative.

We have already seen how this kind of error propagation can happen even without AI. In my own

community, Owens Valley Paiute learners have begun using the word for “come” in metaphorical ways, as in the English phrase “I come from Bishop.” Traditionally, though, the word refers only to physical movement. While the shift may seem minor, it introduces a grammatical change that alters the linguistic heritage we’re trying to preserve. Despite fluent speakers noting that the usage is incorrect, it has continued to gain traction, even among educators. This didn’t happen because of a translation tool or AI system. It emerged gradually through informal teaching and second-language learning, where English-influenced patterns crept in over time. But if AI tools were to reinforce these same patterns, especially in the absence of community validation, they could dramatically accelerate such shifts. What currently unfolds over years through word-of-mouth could happen much faster through the widespread use of flawed models. Indeed, some communities have raised concerns about unvalidated AI-generated language resources. For example, Anishinaabemowin speakers noted the circulation of materials created using large language models that lacked community review or verification [14]. This is the double bind of endangered language technology: the less data we have, the harder it is to build accurate systems, but also, the higher the cost of getting things wrong. A single mistranslation or structural error, if uncorrected, can echo across generations of learners.

### **Barriers To Ethical Collaboration**

Even well-intentioned researchers often face uncertainty about how to engage responsibly. Funding models and institutional timelines are not always aligned with the relationship-driven nature of Indigenous language work. For example, current requirements such as letters of support from tribal governments are intended to promote transparency and community involvement. At the same time, many impactful language efforts take shape outside of formal tribal structures but through grassroots programs, individual teachers, or local initiatives that may not be directly connected to tribal governance. These efforts often have deep community trust and on-the-ground experience, but may face challenges navigating institutional processes. Supporting these contributors and helping them participate fully in research collaborations requires an understanding of how authority, consent, and accountability function in community-led language work.

These challenges are compounded by the fact that many Indigenous communities have long and often painful histories with outside researchers. As Linda Tuhiwai Smith has famously noted, the term “research” can evoke past experiences of mistrust in some Indigenous communities [15]. In my own, earlier documentation efforts have sometimes been received with suspicion or regret. For AI researchers entering this space, these histories must be understood not as peripheral context, but as essential background that shapes how new work is received. As such, ethical engagement is not only about clearing institutional hurdles. It also involves building trust and understanding lived experiences. Determining who can speak for a community, how consent is negotiated, and what forms of reciprocity are appropriate all require grounded relationships and sustained dialogue.

Some researchers have begun exploring what it looks like to build AI-powered tools for Indigenous languages in ways that are grounded in community engagement. Santos Pinhanez et al. [16] frame collaboration with Indigenous communities as both a means of enhancing research quality and a valuable opportunity to explore longstanding AI challenges such as explainability, common-sense reasoning, and contextual modeling. At the same time, Bird [17] offers an important caution: the scientific opportunities that arise from this work must not eclipse its purpose. The goal of speech and language technology should not be to apply AI to Indigenous contexts for its own sake, but to meet community needs as defined by those communities. When done well, this work may indeed surface compelling technical problems, but the direction must be set locally.

These tensions (e.g., between institutional requirements and working relationships, technical interest and ethical responsibility, scientific curiosity and community-defined goals, etc.) are precisely the kinds of questions our workshop aims to surface. By creating space for open, grounded

discussion among AI researchers, language practitioners, and community members, we hope to support more thoughtful, durable approaches to collaboration in this fast-evolving space.

## Proposal: A Workshop On AI And Indigenous Language Revitalization

We propose a three-day workshop to be held in May 2026 at Loyola Marymount University in Los Angeles, California. This convening will bring together Indigenous language teachers, linguists, archivists, and AI researchers to collectively shape the future of AI in endangered language revitalization contexts. The workshop will focus on surfacing shared concerns, identifying best practices, and collaboratively drafting concrete recommendations for ethical, community-centered AI engagement. By grounding these conversations in both lived experience and technical expertise, we aim to create a durable foundation for interdisciplinary collaboration that centers community priorities.

The workshop will be organized and facilitated by an interdisciplinary team with strong community ties, technical expertise, and experience in data and AI infrastructure:

- **Jared Coleman** (Chair and Principal Investigator) is an Assistant Professor of Computer Science at Loyola Marymount University and a member of the Big Pine Paiute Tribe of the Owens Valley. His research explores how large language models can be used for extremely low-resource language revitalization efforts. He brings a dual perspective as both a technical researcher and community member, helping to bridge AI development with the ethical and cultural considerations of Indigenous language revitalization.
- **Bhaskar Krishnamachari** is a professor of Electrical Engineering, Computer Engineering, and Computer Science at the University of Southern California. A widely recognized expert in artificial intelligence, he brings broad technical expertise to the development of innovative, responsible AI systems. He collaborates closely with Dr. Jared Coleman on LLM-based methods for extremely low-resource languages, contributing deep knowledge of model design and evaluation in settings with limited data.
- **Tainā Coleman** is a postdoctoral researcher at the San Diego Supercomputer Center. She works on the NSF-supported *National Data Platform (NDP)*, a federated and extensible data ecosystem designed to support cross-disciplinary collaboration [18]. She brings expertise in data infrastructure and is particularly interested in how the NDP can be adapted to support AI digital humanities projects, especially in contexts that require nuanced treatment of sovereignty, access, and ethical constraints.
- **Glenn Nelson** is a language teacher with the Owens Valley Career Development Center's Nüümü Yadoha Language Program, a member of the Bishop Paiute Tribe, and a long-time community language documenter. As both a carrier and teacher of Owens Valley Paiute, he brings a vital perspective on the stakes of linguistic accuracy in revitalization work. His experience underscores how even small, English-influenced grammatical shifts can undermine linguistic integrity over time, especially when they become codified in teaching materials or reinforced by digital tools. His participation ensures that workshop discussions remain grounded in the lived realities of language education and the nuanced challenges of sustaining linguistic traditions across generations.
- **Khalil Iskarous** is a Professor of Linguistics at the University of Southern California with expertise in laboratory phonology, computational linguistics, and endangered language revitalization. He has taught field courses on endangered language documentation for Atayal,

Saisyat (Taiwan), and Ladin (Italy), and has collaborated closely with Dr. Jared Coleman on AI methods for extremely low-resource languages. Dr. Iskarous also brings experience teaching courses on AI from a linguistic perspective, with a strong emphasis on responsible applications in endangered language contexts.

## Workshop Objectives

This workshop is structured around a core set of objectives that reflect both the urgency of the current moment and the depth of experience held by the communities involved in language revitalization. Drawing on the expertise of Indigenous language teachers, DLI/DEL practitioners, linguists, and AI researchers, the event aims to surface shared concerns, identify promising approaches, and chart a forward-looking agenda for responsible work at the intersection of AI and endangered languages. Participants will work collaboratively toward the following goals:

- Reflect on parallels between early language documentation efforts and current AI practices, particularly where coordination, transparency, or validation may be lacking.
- Surface current and emerging harms, as well as promising practices, in the use of AI for Indigenous languages.
- Explore models for community governance, data stewardship, and co-design that can be adapted or expanded for AI-driven contexts.
- Draft concrete, field-informed recommendations for researchers, institutions, and funding agencies seeking to support this work responsibly.
- Seed a sustainable cross-disciplinary network to support ongoing dialogue, collaboration, and project incubation.

By explicitly centering the knowledge and priorities of communities with deep experience navigating language documentation, data sovereignty, and community-based research, these objectives aim to produce outputs that are practically actionable. The workshop will create the conditions for long-term collaboration and establish shared guidance that can shape future work in both academic and community settings.

## Workshop Format And Structure

The workshop will include a mix of invited talks, cross-disciplinary panels, breakout sessions, and open discussion. These formats are designed to create space for both structured knowledge sharing and relationship-driven collaboration across disciplinary and community boundaries. Key components will include:

- **Invited Talks**, featuring leaders in Indigenous language revitalization, documentation, and AI. These sessions will highlight lessons from both community practice and technical research, and help establish shared context for deeper discussion.
- **Cross-Disciplinary Panels**, where Indigenous educators, linguists, archivists, and AI researchers will engage in moderated dialogue around critical challenges and emerging opportunities in this space. These panels will help surface areas of alignment as well as points of tension that merit further exploration.

- **Cross-Disciplinary Breakout Sessions**, held as working lunches and facilitated in small groups, will provide space for participants from different domains to collaborate directly. These sessions will focus on case-based problem solving, the articulation of shared values, and the development of field-informed guidance.
- **Synthesis and Action Planning**, during which participants will collaboratively draft shared principles, identify concrete recommendations, and outline possible next steps for continued collaboration and knowledge exchange.

To ensure broad accessibility and deep engagement, we will provide travel support for participants. Our goal is not only to produce actionable outputs, but to foster long-term relationships, trust, and a shared foundation for future interdisciplinary work.

### **Participant Recruitment And Support Plan**

We will assemble a balanced group of participants with a range of professional backgrounds, experience, and areas of expertise. In keeping with NSF's commitment to ensuring that opportunities are available to all Americans regardless of protected characteristics, the participant group will reflect a wide range of geographic regions, institutional types, and professional backgrounds. In addition to inviting established researchers and practitioners in AI, linguistics, and language documentation, we will actively seek participation from community-based language workers, early-career scholars, and individuals with direct experience in revitalization efforts. Travel support will be prioritized for participants from under-resourced institutions or regions, and for early-career researchers, language teachers, and others without institutional funding. The workshop will be held in a physically accessible venue and will be designed to support full participation from attendees with a range of communication styles and professional training. This approach is designed to ensure that the conversations and recommendations that emerge from the workshop are informed by a wide spectrum of perspectives, with particular emphasis on those with direct experience in language revitalization work and those who are most affected by the outcomes of AI system design.

### **Family Care And Accessibility Considerations**

We recognize that the ability to participate in events such as this workshop may be influenced by family care responsibilities. To support attendees with caregiving needs, we will provide information about local childcare providers near the conference site, as well as options for short-term care services. We will also include details about nearby family-friendly accommodations and facilities in the participant welcome materials. Attendees will be invited to indicate any specific needs related to family care in their registration form, so that we can assist with identifying appropriate local resources in advance of the event. These efforts are intended to help ensure that participation is feasible for individuals managing a range of personal and family obligations.

### **Participant Conduct And Safety**

In accordance with NSF policy, the organizers will adopt and disseminate a clear code of conduct that addresses sexual harassment, other forms of harassment, and sexual assault. This code will include clear procedures for submitting complaints and a plan for addressing any violations that occur. The policy will be distributed to all participants prior to the workshop and made available onsite.

### **Anticipated Outputs**

The workshop is designed not only to foster discussion, but also to produce concrete and lasting outcomes. We anticipate the following outputs:

- A **position paper** articulating key considerations, stakeholder perspectives, and shared guidance for responsible AI in endangered language contexts.
- A **set of actionable recommendations** for researchers, institutions, and funders, outlining best practices for ethical and effective AI projects in endangered language contexts.
- The formation of an **interdisciplinary working group or mailing list** to support long-term coordination and knowledge exchange among workshop participants.

These outputs will help catalyze ongoing collaboration and ensure that insights from the workshop continue to inform future work in this area.

### **Related Workshops And Distinct Contributions**

A number of recent workshops and initiatives have explored the intersection of Indigenous knowledge systems and artificial intelligence, or have supported the development of Indigenous language technologies more broadly. These efforts offer valuable insights and continue to shape a growing ecosystem of research and practice. However, our proposed workshop addresses a distinct gap: it centers explicitly on the ethical design and implementation of AI research in endangered language contexts, and is intentionally cross-disciplinary and rooted in the practical priorities of Indigenous language teachers and documentation experts who have long worked in ethical collaboration with tribal communities.

For example, *AmericasNLP* [9], the *Indigenous in AI/ML Affinity Workshop* at NeurIPS [8], and *ComputEL* [10] offer vital platforms for technical work in natural language processing for Indigenous languages. Many participants in these venues are Indigenous or collaborate closely with Indigenous communities, yet the focus remains primarily on technical AI/ML research. Discussions around community involvement and long-term reciprocity are typically limited or secondary. By contrast, gatherings such as the *DLI Community of Science* [13] (supported by the NSF DLI/DEL program) foreground collaborative research infrastructure, community governance, and sustainable models for language documentation. This community offers strong models for ethical engagement and interdisciplinary practice, though they have not directly addressed the emerging challenges and risks posed by AI technologies. The *Indigenous Protocol and Artificial Intelligence Workshops* [19] and newer initiatives such as *Abundant Intelligences* [20] and the *ENRICH* [21] project reflect growing interest in Indigenous-led approaches to AI. Their primary contributions lie in conceptual frameworks, arts-based inquiry, and digital sovereignty, rather than in applied language technologies or direct collaboration with community-based language revitalization practitioners. The *Advancing Indigenous Language Technologies (AILT)* [22] working group (also supported by an NSF DLI/DEL award) brings together academic and community researchers to build AI-driven tools for Indigenous languages. Their work in areas such as speech recognition, predictive text, and digital resource development illustrates the increasing technical capacity for community-aligned language technology projects. Like other initiatives mentioned above, AILT contributes to a rapidly evolving landscape that our proposed workshop seeks to complement by creating space for cross-disciplinary exchange and practical, field-driven guidance.

What sets our proposed event apart is its emphasis on generating actionable guidance for those working at the intersection of AI and endangered languages. Rather than centering technical development or collaboration frameworks alone, the workshop is designed to leverage the lived expertise of Indigenous language teachers and the applied experience of the DLI/DEL community that has spent decades building effective models for working with Indigenous communities. By bringing these voices into conversation with AI researchers, the workshop aims to co-develop concrete recommendations for practices that reflect the priorities of Indigenous communities and the

realities of language revitalization work. This will be the first time this workshop is held. To our knowledge, no prior convening has brought together this particular combination of stakeholders to collectively define principles and practices for responsible, community-centered AI in the context of Indigenous language revitalization.

## Intellectual Merit

This project will generate new conceptual and methodological frameworks at the intersection of artificial intelligence, linguistics, and language documentation by addressing how computational tools can be responsibly adapted for endangered language contexts. These settings present unique technical challenges such as data scarcity, grammatical/morphological complexity, and limited access to validation resources that are under-explored in current AI research. The workshop will bring together experts in language documentation, linguistics, and AI to surface shared research challenges and identify strategies for aligning AI-powered tools with linguistic integrity and locally articulated needs. In doing so, it will contribute to foundational research in artificial intelligence by advancing methods for low-resource language model adaptation, human-AI collaboration, and responsible system design—areas that align directly with NSF’s strategic priorities in AI. The insights produced will help lay the groundwork for future inquiry that would not be possible without improved trust, infrastructure, and communication across fields. The resulting recommendations will guide both ethical and technical best practices, enabling research that is not only socially grounded but also methodologically rigorous and better equipped to address the linguistic complexity and data constraints endemic to this domain.

## Broader Impacts

The proposed workshop will support the development of scientifically informed, community-centered approaches to AI and Indigenous language revitalization. By convening language teachers, linguists, archivists, and AI researchers, the event will generate actionable recommendations for responsible research design and collaborative development practices. These outputs, along with a position paper and an interdisciplinary working group, will serve as durable resources for both academic and community stakeholders. In the near term, the project will help communities make informed decisions about AI integration, while improving the usability and accountability of language technologies. In the longer term, it will lower structural barriers that currently inhibit scientifically meaningful work in this area, such as limited trust, opaque data practices, and unclear evaluation standards. By building relationships across disciplines, the workshop will lay the groundwork for new lines of inquiry into under-explored questions involving Indigenous languages, enabling more ambitious and methodologically rigorous research that is responsive to community priorities. It will also create early entry points for emerging scholars in both AI and language revitalization, helping them develop the relationships, knowledge, and guidance needed to pursue impactful work in this evolving field.

## Archival Plan

This proposal does not involve the creation or collection of new linguistic data.

## Results From Prior NSF Support

**NSF Award Number:** 2451267

**PI:** Dr. Jared Coleman

**Amount:** \$150,000

**Period of Support:** 08/01/2025 – 07/31/2027

**Title:** CRII: CSR: RUI: *Novel Approaches for Task Graph Scheduling Algorithm Design, Evaluation, and Comparison*

Although this award was recently made and work has not yet begun (project start date: August 1, 2025), we summarize below the major goals and anticipated impacts of the funded project.

**Intellectual Merit:** This project aims to advance the understanding of task graph scheduling algorithms in distributed computing environments by developing a principled, reproducible evaluation framework. It builds on prior work demonstrating that commonly used benchmarking approaches can obscure performance failures on adversarial instances. The project will extend an open-source software framework (SAGA) to support new problem variants (e.g., constraint-based and stochastic scheduling) and integrate it with real-world distributed computing platforms. The research will also explore hybrid algorithm design and improvements to machine learning-based scheduling methods. These contributions will inform the design of more robust, efficient, and context-sensitive scheduling algorithms.

**Broader Impacts:** The project will lower barriers to reproducible evaluation in distributed systems research and provide an extensible platform for future studies. It will also support undergraduate research at a primarily undergraduate institution, broaden participation in computing, and result in publicly available tools and datasets. Findings from the project are expected to have broad utility for researchers and practitioners working on scientific workflows, IoT systems, and distributed AI pipelines.

**Publications:** No publications have yet been produced under this award.

**Research Products and Availability:** No research products have yet been generated. All code and data developed under this award will be made available through public repositories associated with the SAGA project.

**Relation to Proposed Work:** While the funded project focuses on algorithm evaluation and design for distributed task scheduling in computational systems, the current proposal addresses the ethical, community-centered integration of AI technologies into Indigenous language revitalization contexts. There is no technical or effort overlap between the two projects. However, both projects reflect a broader commitment to responsible AI development, and share methodological values such as transparency, interdisciplinary collaboration, and contextual evaluation of algorithmic tools.

## References Cited

- [1] J. Coleman, B. Krishnamachari, R. Rosales, and K. Iskarous, “LLM-assisted rule based machine translation for low/no-resource languages,” in *Proceedings of the 4th Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2024)*, M. Mager, A. Ebrahimi, S. Rijhwani, A. Oncevay, L. Chiruzzo, R. Pugh, and K. von der Wense, Eds. Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 67–87. [Online]. Available: <https://aclanthology.org/2024.americasnlp-1.9/>
- [2] C. Anderson, M. Nguyen, and R. Coto-Solano, “Unsupervised, semi-supervised and LLM-based morphological segmentation for Bribri,” in *Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)*, M. Mager, A. Ebrahimi, R. Pugh, S. Rijhwani, K. Von Der Wense, L. Chiruzzo, R. Coto-Solano, and A. Oncevay, Eds. Albuquerque, New Mexico: Association for Computational Linguistics, May 2025, pp. 63–76. [Online]. Available: <https://aclanthology.org/2025.americasnlp-1.7/>
- [3] G. Tanzer, M. Suzgun, E. Visser, D. Jurafsky, and L. Melas-Kyriazi, “A benchmark for learning to translate a new language from one grammar book,” in *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net, 2024.
- [4] C. Zhang, X. Liu, J. Lin, and Y. Feng, “Teaching large language models an unseen language on the fly,” *CoRR*, vol. abs/2402.19167, 2024.
- [5] R. Jimerson, Z. Liu, and E. Prud’hommeaux, “An (unhelpful) guide to selecting the best ASR architecture for your under-resourced language,” in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 1008–1016. [Online]. Available: <https://aclanthology.org/2023.acl-short.87/>
- [6] E. Le Ferrand, Z. Liu, A. Arppe, and E. Prud’hommeaux, “Are modern neural ASR architectures robust for polysynthetic languages?” in *Findings of the Association for Computational Linguistics: EMNLP 2024*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 2953–2963. [Online]. Available: <https://aclanthology.org/2024.findings-emnlp.166/>
- [7] Z. Liu, C. Richardson, R. J. Hatcher, and E. Prud’hommeaux, “Not always about you: Prioritizing community needs when developing endangered language technology,” in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, S. Muresan, P. Nakov, and A. Villavicencio, Eds. Association for Computational Linguistics, 2022, pp. 3933–3944. [Online]. Available: <https://doi.org/10.18653/v1/2022.acl-long.272>
- [8] Indigenous in AI, “Indigenous in ai at neurips 2024,” <https://neurips.cc/virtual/2024/affinity-workshop/50035>, 2024, affinity Workshop at NeurIPS, Accessed July 2025.
- [9] AmericasNLP Workshop Organizers, “Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas,” <https://turing.iimas.unam.mx/americasnlp/>, 2025, accessed July 2025.

- [10] ComputeL Workshop Organizers, “ComputeL: Workshop on Computational Methods for Language Documentation and Revitalization,” <https://computel-workshop.org/>, 2025, accessed July 2025.
- [11] S. D. Penfield, A. Serratos, B. V. Tucker, A. Flores, O. Harper, and N. Hill, “Community collaborations: Best practices for north american indigenous language documentation,” *International Journal of the Sociology of Language*, vol. 191, pp. 187–202, 2008.
- [12] S. R. Carroll, I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera, M. Parsons, K. Raseroka, D. Rodriguez-Lonebear, R. Rowe, R. Sara, J. D. Walker, J. Anderson, and M. Hudson, “The care principles for indigenous data governance,” *Data Science Journal*, Nov 2020.
- [13] K. Hildebrandt and R. Sapien, “Establishing a dynamic language infrastructure community of science,” National Science Foundation Award Abstract #2301800, 2023, accessed: 2025-07-19. [Online]. Available: [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=2301800](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2301800)
- [14] M. Becking, “Fraudulent anishinaabemowin resources a serious concern,” *Anishinabek News*, October 2024, accessed: 2025-07-19. [Online]. Available: <https://anishinabeknews.ca/2024/10/22/fraudulent-anishinaabemowin-resources-a-serious-concern/>
- [15] L. T. Smith, *Decolonizing Methodologies: Research and Indigenous Peoples*, 3rd ed. London: Zed Books, 2021, oCLC 1242806586.
- [16] C. S. Pinhanez, P. Cavalin, M. Vasconcelos, and J. Nogima, “Balancing social impact, opportunities, and ethical constraints of using ai in the documentation and vitalization of indigenous languages,” in *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23*, E. Elkind, Ed. International Joint Conferences on Artificial Intelligence Organization, 8 2023, pp. 6174–6182, aI for Good. [Online]. Available: <https://doi.org/10.24963/ijcai.2023/685>
- [17] S. Bird, “Decolonising speech and language technology,” in *Proceedings of the 28th International Conference on Computational Linguistics*, D. Scott, N. Bel, and C. Zong, Eds. Barcelona, Spain (Online): International Committee on Computational Linguistics, Dec. 2020, pp. 3504–3519. [Online]. Available: <https://aclanthology.org/2020.coling-main.313/>
- [18] I. Altintas, C. Meertens, A. Gupta, I. Rodero, and M. Floca, “National data platform pilot: Services for equitable open access to data,” National Science Foundation Award Abstract #2333609, 2023, office of Advanced Cyberinfrastructure (OAC), University of California San Diego. Start Date: September 1, 2023. End Date: August 31, 2026. Accessed: 2025-07-19. [Online]. Available: [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=2333609](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2333609)
- [19] J. E. Lewis, A. Abdilla, N. Arista, K. Baker, S. Benesiinaabandan, M. Brown, M. Cheung, M. Coleman, A. Cordes, J. Davison, K. Duncan, S. Garzon, D. F. Harrell, P.-L. Jones, K. Kealiikanakaoleohaillani, M. Kelleher, S. Kite, O. Lagon, J. Leigh, M. Levesque, K. Mahelona, C. Moses, I. I. Nahuewai, K. Noe, D. Olson, 'Owi Parker Jones, C. R. Wolf, M. R. Wolf, M. Silva, S. Fragnito, and H. Whaanga, “Indigenous protocol and artificial intelligence position paper,” Indigenous Protocol and Artificial Intelligence Working Group and the Canadian Institute for Advanced Research, Honolulu, HI, Project Report, 2020, edited by Jason Edward Lewis. Submitted version. Accessed 2025-07-19. [Online]. Available: <https://spectrum.library.concordia.ca/id/eprint/986506/>

- [20] J. E. Lewis, H. Whaanga, and C. Yolgörmez, “Abundant intelligences: placing ai within indigenous knowledge frameworks,” *AI & Society*, vol. 40, no. 4, pp. 2141–2157, 2025. [Online]. Available: <https://doi.org/10.1007/s00146-024-02099-4>
- [21] ENRICH Coordinating Hub, “About — enrich: Equity for indigenous research and innovation,” Online, 2025, accessed: 2025-07-19. [Online]. Available: <https://www.enrich-hub.org/about>
- [22] A. Fountain, E. Jackson, and A. Tsosie-Paddock, “Collaborative research: Advancing indigenous language technologies,” NSF Award #2347147, 2024, division of Behavioral and Cognitive Sciences, University of Arizona. Start Date: September 15, 2024. End Date: August 31, 2025. [Online]. Available: [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=2347147](https://www.nsf.gov/awardsearch/showAward?AWD_ID=2347147)