



AI Policy Must Respect Indigenous Data Sovereignty

Kerri J. Malloy, San José State University

As artificial intelligence spreads across public and private life, U.S. policymakers face growing pressure to govern it, but **regulation remains uneven**. In that unsettled landscape, any AI framework that treats Tribal Nations as just another stakeholder group starts from the wrong premise. When AI systems rely on data about Tribal Nations, American Indian and Alaska Native communities, or Native Hawaiian communities, the issue is not simply privacy, bias, or access, but authority. AI governance that ignores **Indigenous Data Sovereignty** reproduces extractive practices in digital form and denies Tribal Nations authority to shape AI around their own priorities and knowledge systems.

Responsible AI policy requires more than general commitments to fairness or inclusion. Policymakers should require Indigenous Data Sovereignty plans, mandate early and meaningful tribal consultation, tie public funding and approvals to clear data-governance standards, and impose real consequences when agencies, vendors, or research partners use Native data without proper tribal authority.

Sovereignty Is Not Privacy

For **federally recognized tribes**, the U.S. framework is **government-to-government**. Tribal Nations are **sovereign** political entities with a legal relationship to the federal government, so data about Tribal citizenship, lands, health, education, language, culture, and community life cannot be treated as ordinary administrative information. These are collective rights tied to Tribal governance and Indigenous self-determination. Treating such data as ordinary inputs weakens the policy response from the start.

Indigenous Data Sovereignty means Tribes **govern** how data about their communities are collected, used, stored, interpreted, shared, and applied. Access alone does not give a university, agency, or contractor the right to train AI systems on Native data.

Native Data Is More Than Datasets

AI intensifies an existing **problem**. Large technical systems are often built from legacy records, public datasets, archives, cultural materials, and research collections assembled for purposes far removed from AI. Once materials are labeled public or research-ready, institutions often treat reuse as automatic. That is a governance failure. For Native communities, data is not limited to formal datasets; it also includes language, traditional knowledge, photographs, recordings, and archival materials long treated as available for reuse. Much of that material was collected or preserved without meaningful prior consent, making later AI reuse an extension of older extractive practices. Collective Benefit, Authority to Control, Responsibility, and Ethics (**CARE**) and Ownership, Control, Access, Possession (**OCAP**) principles for Indigenous data sovereignty make the same point: data governance is about who decides, who benefits, and who has the power to say no.

The representational problem is just as serious. Data used to make claims about Native communities are often too **partial** or too small to represent the diversity of Tribal Nations and Native communities. AI systems can turn those limited inputs into broad outputs that look authoritative. In practice, that can flatten distinctions among Tribal Nations, strip information from its cultural and political context, and give weak data an undeserved aura of authority. The issue is not only inclusion in datasets, but control over how Native realities are categorized, interpreted, and built into technical systems.

Tribes Must Shape AI Early

Tribal consultation is widely recognized across public policy, but too often it still happens after key decisions have been shaped, functions as a procedural comment period rather than government-to-government engagement and leaves Tribal Nations with little evidence that their participation changed the outcome. If agencies wait until a model is trained, a vendor is selected, or procurement is nearly complete, consultation has lost much of its value. In the AI context, **meaningful consultation** must begin before design and procurement decisions are locked in, require agencies to document how Tribal input shaped the outcome, and create a path for Tribal Nations to help govern how systems are designed, evaluated, and used.

Policymakers should apply that standard across AI governance. Tribes should be involved before data are acquired, licensed, or used for training; before grants and contracts are awarded; before agencies buy AI tools; and before systems are deployed in public programs or research. Consultation after design is only notice. The stronger goal is governance: Native communities should help shape what systems are built, what purposes they serve, and what limits govern their use.

What Policymakers Should Do to Protect Indigenous Data Sovereignty

Closing that gap requires enforceable standards. Four actions are especially important.

First, any publicly funded or deployed AI system that uses Native data or significantly affects Tribal Nations should be governed by an Indigenous Data Sovereignty plan developed through meaningful consultation. That plan should specify permitted uses, access, retention, downstream sharing, commercialization limits, and rules for vendor access, partnerships, memoranda of understanding, and external contracts.

Second, Tribal consultation must occur early enough to shape outcomes. Agencies should consult before major design, contracting, and procurement decisions; provide enough information for Tribal Nations to assess proposals; and explain how data will be used, shared, retained, and governed over time.

Third, public funding should reinforce these requirements rather than bypass them. Grants, contracts, and research approvals should move forward only when enforceable data-use terms are in place and CARE and OCAP standards are built into project design. Policymakers should also review funding and data-sharing rules that pressure Tribes or researchers to make Native data public in ways that weaken Tribal governance.

Finally, these requirements must carry real consequences. Agencies, contractors, research institutions, outside vendors, and research partners that ignore consultation obligations, violate agreed data terms, or use Native data without proper governance should face funding consequences, procurement disqualification, or research sanctions. Otherwise, AI policy will reproduce the symbolic practices it claims to correct.

AI policy will not be ethical, accurate, or legitimate if it treats Tribal Nations as data sources rather than governments with authority. The task is to prevent digital extraction and ensure Native communities can shape AI around their own priorities and knowledge systems. If Tribal Nations do not have a governing role, the system is not ready to build, buy, or deploy.