Check for updates

Commentary

BIG DATA & SOCIETY

Big Data & Society
October–December: 1–6
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20539517261498673
journals.sagepub.com/home/bds

Sage

AI is locking out Indigenous voices: Toward more equitable futures

Inti Josue Cachipuendo-Contero[1,2] **and Justin Yeager**[1]

Abstract

Artificial intelligence (AI) increasingly helps communities make decisions related to healthcare, agriculture, education, and climate adaptation, yet most AI systems are built for people in connected, urban settings who speak dominant languages such as English. Indigenous and rural communities are often left out of the data, the model design, and the subsequent decisions that follow: a problem AI has not created, but has intensified. Using the Andes of Ecuador as a case in point, this commentary argues that this is a cyclical problem that is older than AI: where outside actors introduce a technology and decide how it will work without considering diverse cultures of people, often rural or Indigenous, who will use it. Unlike infrastructure projects such as a road or clinic which are commonly built in rural communities without prior consultation, commercial AI systems similarly are currently impacting communities but are controlled by distant companies, leaving community members with no clear local mechanism to contest or correct culturally inappropriate model outputs. Drawing on examples from Kichwa communities, we show how language, timing, and collective decision-making rarely fit the way most AI systems are designed, leading to advice that is misread, mistimed, or simply ignored. We then outline what a negotiated, culturally sensitive approach could look like: outputs checked by local community hierarchy before they reach the broader community, data used only with permission, and voice notes or images accepted as input, not only typed text. This is a concept, offered as a starting point to open a broader discussion in collaboration with developers and policymakers who want AI to serve these communities rather than sideline them. What is at stake is not only access to new tools, but who gets to decide how knowledge is used, and who can refuse that use altogether.

Keywords

Indigenous knowledge systems, AI governance, algorithmic exclusion, epistemic justice, Kichwa language, code-switching

Introduction

Artificial intelligence (AI) tools now shape many parts of daily life, improving efficiency and saving time and effort. Yet most are built for people in connected cities who speak global languages and follow Western ideas about time, logic, and communication. Indigenous and rural communities are often missing from the data, while their speech patterns, planting and ritual calendars, local governance, and limited internet access are routinely ignored. These omissions matter not only for access to AI but also because AI-generated recommendations increasingly inform community decisions about healthcare, education, disaster response, agricultural support, and communal resource allocation. If Indigenous ways of knowing are excluded from the start, AI risks repeating older colonial patterns in which outside actors or institutions define problems, categories, and solutions affecting Indigenous communities without meaningful local participation. AI is already used in these communities, but implementation remains relatively new, leaving time for meaningful change.

This commentary grows from ongoing dialogue with Indigenous communities across Ecuador and from our experiences. Inti Josue Cachipuendo-Contero is Kichwa

[1]Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Facultad de Ingenierías y Ciencias Aplicadas, Universidad de Las Américas, Quito, Ecuador
[2]Department of Advertising, Public Relations and Audiovisual Communication, Universitat Autònoma de Barcelona, Barcelona, Spain

Corresponding author:
Justin Yeager, Grupo de Investigación en Biodiversidad, Medio Ambiente y Salud (BIOMAS), Facultad de Ingenierías y Ciencias Aplicadas, Universidad de Las Américas, Quito, Ecuador.
Email: yeagerjd@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (https://us.sagepub.com/en-us/nam/open-access-at-sage).

and has lived in Andean communities throughout his life, participating in community decisions and collective planning. Justin Yeager has spent over a decade conducting intercultural research with Indigenous groups. We also draw on conversations with a Waorani university student about the growing use of mobile phones and AI among young people and with a Kichwa university professor with decades of experience in higher education and Indigenous governance. These relationships developed through longstanding academic ties to Indigenous communities in Ecuador, including residence in communities and sustained engagement, and informed the examples and conceptual model developed here. We do not claim to speak for all Indigenous peoples; our discussion draws on these experiences and those of colleagues and peers.

How and why are Indigenous communities left out of AI?

The problem extends beyond internet access, antiquated devices, or missing Indigenous-language text and speech data. The deeper issue is a systemic mismatch. AI systems typically expect standard text, fixed calendar time, and an individual making a decision, while Indigenous knowledge and cultures may operate differently. Spanish dominates in Ecuador, but many Indigenous territories maintain their own languages, and meaning can shift between Kichwa and Spanish within a single sentence. Some decisions, from a *minga* (community-led service project) to repair a road or prepare an event to choosing local leadership, pass through assemblies involving elders or local leaders because they affect the broader community. Thus, we must move beyond "inclusive AI" when it means only better access or translation.

AI instead recapitulates an older pattern in which outside actors introduce systems or interventions without meaningful consultation with those expected to use them. Indigenous communities have faced this, for example, with health programs designed elsewhere that overlook local languages, practices, or trusted forms of care. Such problems can often be anticipated, mitigated, or corrected through consultation with a local agency or provider. Commercial AI is harder to contest because the companies designing and operating models are geographically and culturally distant, often leaving communities without a local representative when problems arise. This commentary therefore asks what meaningful negotiation and direct collaboration between rural and Indigenous communities and AI developers would require.

In many Indigenous communities, knowledge is produced and shared collectively through stories, ritual practice, and communal deliberation rather than held privately by individuals. Elders and local leaders may help determine how different information should circulate. Within community governance, information about assemblies, collective decisions, or community projects may circulate broadly, while details concerning family welfare, internal disputes, or conflict-resolution processes may remain limited to those involved or authorized to review them. This differs from AI systems trained on vast amounts of text aggregated across individual users rather than information collectively agreed upon by a community. Recent work therefore argues that AI should be situated within Indigenous knowledge frameworks rather than a single global template (Lewis et al., 2025; Perera et al., 2025), including community involvement in classifying knowledge, controlling data access and consent, and supporting ancestral language revitalization. These are essential advances, but they say less about mixed-language speech, ceremonial timing, and collective authority in everyday use, factors affecting whether AI outputs are intelligible, timely, and socially legitimate. We develop these issues through Andean examples.

Our goal is to identify practical considerations for developers, researchers, and policymakers seeking AI systems that can work with Indigenous and other underrepresented rural communities. Collaborative consultation can help prevent stories, rituals, and land-based knowledge from becoming decontextualized data or imprecise outputs that erase who may share knowledge, with whom, and under what conditions.

Unequal by design

AI can guide planting using weather and soil data, predict threats such as floods and landslides (Albahri et al., 2024), and support AI-assisted screening in Indigenous healthcare settings (Perera et al., 2025).

Yet these tools often assume steady internet, affordable data, fluency in English or formal Spanish, and one person asking one question and receiving one answer. In Indigenous contexts, these assumptions shape what AI can understand and how it responds. They perpetuate cognitive colonialism by treating ways of knowing such as written text as the default and oral or communal knowledge as exceptions (Lewis et al., 2025).

This pattern is built into how large language models work. Models learn patterns from vast quantities of digitized text, so performance in a language depends heavily on its online footprint. As Figure 1 shows, Quechua is a macro-language family spoken by an estimated 8–10 million people across the Andes, with Kichwa as Ecuador's branch (Andrade Ciudad and Howard, 2021; Cerrón-Palomino, 1987). In Ecuador, the Instituto Nacional de Estadística y Censos (INEC, 2023) reports that 1.3 million of 16.9 million people self-identify as Indigenous and 527,333 speak Kichwa. Despite this demographic presence, Kichwa had no dedicated computational-language resources until 2024, when its first speech dataset

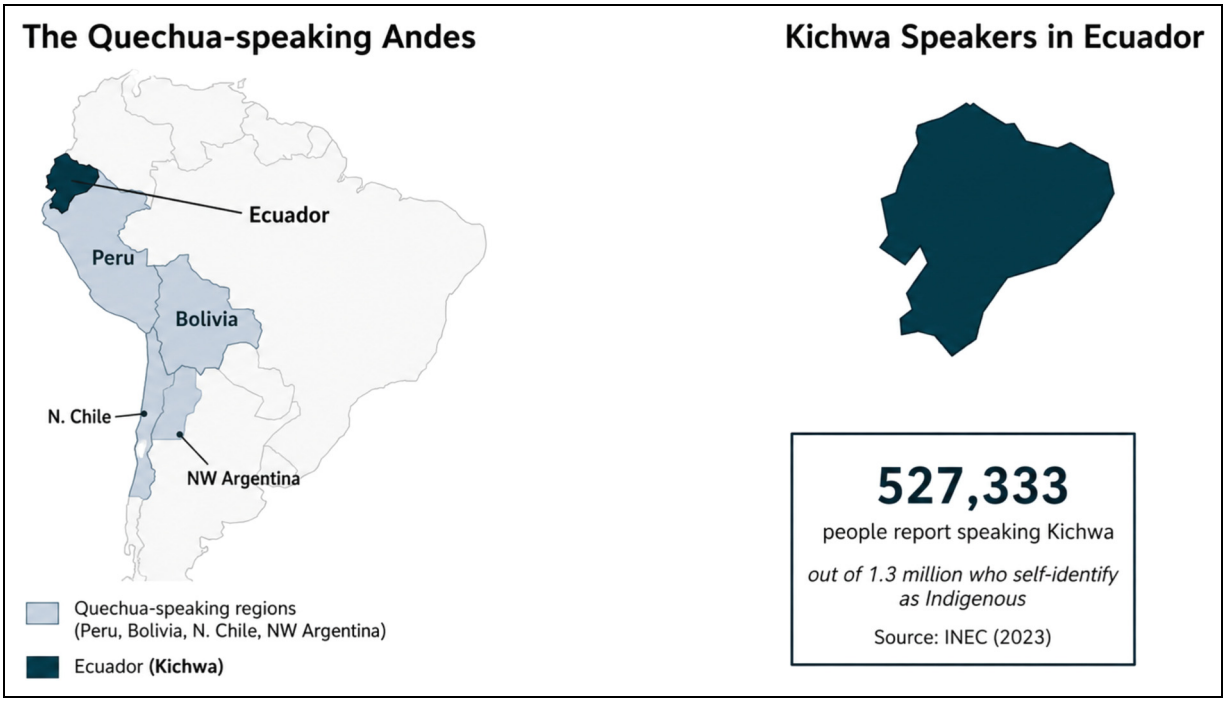


Figure 1. Regional distribution of the Quechua macro-language family across the Andes and reported Kichwa-speaking population in Ecuador. Source: Authors' visualization based on Andrade Ciudad and Howard (2021), Cerrón-Palomino (1987), and Instituto Nacional de Estadística y Censos (2023).

was built from a single radio program (Taguchi et al., 2024). Yet this is not solely a problem of data volume: statistical language modelling can reproduce linguistic patterns without necessarily capturing the relational, temporal, or collective conditions that give those expressions meaning.

AI improvement is further constrained because communities often lack channels to report or correct culturally specific failures directly with AI companies. Such problems may therefore go unnoticed. Katz et al. (2024) report that in 2019, 45.29% of Indigenous people in Ecuador had an active mobile phone, while slow internet and expensive data contributed to near-universal WhatsApp use. A WhatsApp-based AI tool could therefore reach users at a scale that a new stand-alone application may not.

The epistemic divide

A WhatsApp-based tool alone does not resolve this deeper mismatch. Most systems take input from one user and return one answer, rather than a response that may need to reflect a collective decision. Indigenous knowledge, however, is often oral, ritual, and cyclical, transmitted through stories and spoken teaching, tied to ceremony, and organized around solar and seasonal rhythms rather than fixed dates. The Abundant Intelligences framework therefore asks us to imagine not one universal AI model but multiple systems able to engage different ways of knowing (Lewis et al., 2025), including voice input, mixed-language recognition, and routing sensitive information through group rather than only individual channels.

In Andean communities, knowledge is also lived through seasons and land. Knowing when to plant is not simply a calendar date; it depends on rainfall, soil, the moon, and collective decision-making. In Nizag, a Kichwa community in Chimborazo, plant knowledge is tied to places along part of the *Qhapaq Ñan*, the old Andean road system: rather than extracting a fact for independent use, learning requires walking the path with someone who knows it (Chávez Velásquez et al., 2024).

AI must therefore engage more than an added layer of "cultural context." Andean planting may follow *Killa Raymi*, the moon festival, and *paridad*, or balance, through which corn, beans, and squash are grown together not only for yield but to sustain spiritual and ecological balance. When seeking planting information, a community member therefore needs more than word-for-word translation: timing and meaning arise from place, season, and community together.

Language and misrecognition

Everyday Indigenous language further demonstrates the limits of translation accuracy. A Kichwa-speaking farmer might ask, "*quiero purutu para julio, ¿haykaq debo sembrar?*," mixing Kichwa and Spanish in a way fully intelligible and commonly used locally. *Purutu* means beans

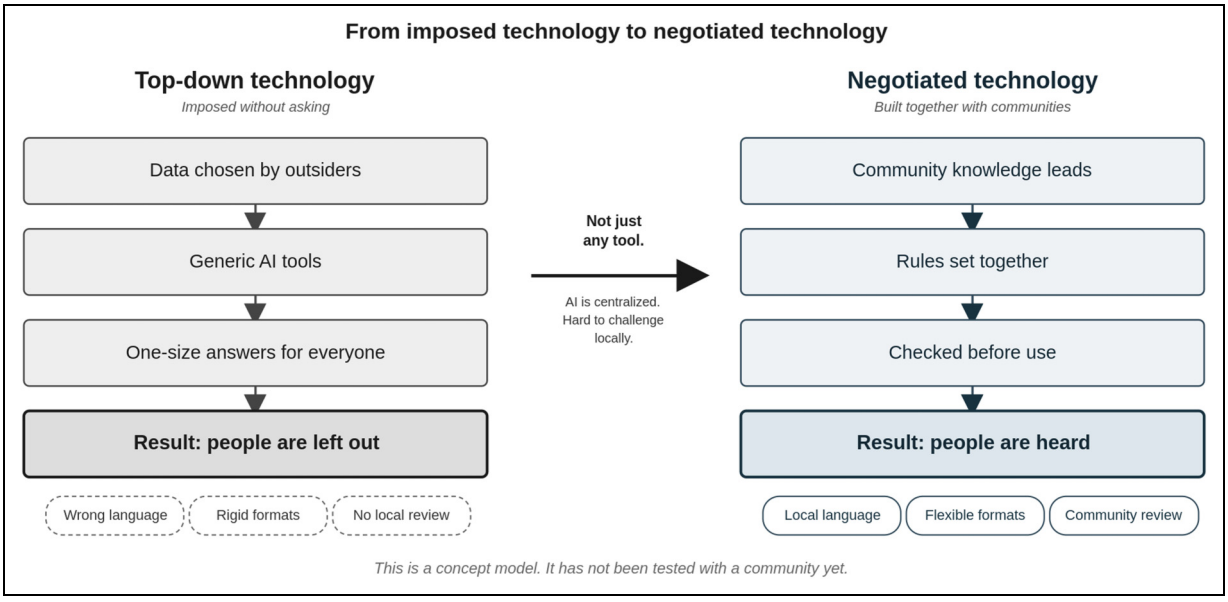


Figure 2. From imposed to negotiated technology. The left panel shows externally selected data and artificial intelligence (AI) tools producing generic outputs that exclude communities; the right shows an alternative based on community knowledge, jointly agreed rules, and review before use. The center highlights AI's particular accountability challenge: commercial models may have no local representative when problems arise. This is a conceptual model, not yet tested through structured co-design with a specific community.

and *haykaq* roughly "when" or "how"; the farmer is asking when and how to plant beans for a July harvest. The difficulty is not simply vocabulary: the question carries seasonal, communal, and spiritual meaning that literal translation cannot capture.

AI trained only on standard Spanish or English may flag *purutu* or *haykaq* as errors or interpret the sentence as broken Spanish. Locally, however, this is normal, deliberate code-switching. AI operating in these settings must therefore interpret meaning from context rather than correct it away.

The problem runs deeper than words. Oral traditions transmit knowledge through metaphor, story, gesture, and spiritual meaning. Consider "*hay que sembrar cuando la luna duerme*" ("we must plant when the moon sleeps"): this refers not simply to a lunar-calendar date but to an obligation to plant at the proper moment in relation to land, community, and the wider cosmic order. Research on Indigenous languages describes such meaning through symbolization: literal translation may reproduce the words while missing their timing, implicit ethics, and fuller meaning (Gqweta, 2025). Similarly, processing Indigenous-language archives involves preserving collective and lived meanings that lack fixed dictionary equivalents (Usmanova et al., 2025). Until AI can engage such meanings, it will remain a poor listener in communities that speak, think, and remember differently from what current models recognize. Whether it can adapt at all remains an open question.

Toward community-validated AI engagement

Better bandwidth and language tools remain insufficient if Indigenous governance is added only after a system is built. From the outset, it should shape what data is collected, who reviews outputs, and which trusted channels carry information forward, as outlined in Figure 2. This raises a harder question: can some Indigenous ways of knowing and governing fit within an algorithmic system at all? Inclusion may require negotiating boundaries around what should become data in the first place. Testing the model with communities is a critical next step; its present basis includes processes witnessed within Kichwa community decision-making and consultation with an Andean professor who has observed governance across these communities for decades.

The question, then, is not simply whether Indigenous knowledge can be "included" in AI, but how information is filtered, checked, and shared. AI systems often move data continuously with little human review or community oversight, while Indigenous settings may require different speeds and protocols. Routine agricultural information, such as weather forecasts or public pest alerts, may circulate immediately. A recommendation affecting communal resources such as irrigation water or community farmland, however, may require review by local leaders or an assembly because it affects multiple households. The goal is not to block information but to preserve the social context

governing action. A technically sound recommendation to redistribute irrigation water may still be inappropriate if it ignores community agreements about who has access, when, and under what conditions.

In Ecuador, AI governance must therefore go beyond consent forms and respect Indigenous peoples' right to govern their knowledge and territory under the country's plurinational constitution. Communities should help determine how local knowledge is encoded and presented in AI systems and outputs, what context accompanies it, who may authorize its use, and how technologies fit existing authority structures. Consent means more than clicking "agree": a household survey may require individual consent, whereas mapping a shared irrigation system may require community agreement. Land, ecology, ritual, and risk carry social meanings that communities, rather than algorithms, are positioned to weigh (Requelme et al., 2024). Such governance is not a barrier to access, but a condition for outputs communities can trust, act upon, and adopt.

Determining what information should circulate freely is perhaps the hardest part and cannot reasonably be left to programmers alone. This does not require individualized approval for every output; rather, access can vary by context, with routine information circulating directly and sensitive cases such as land disputes flagged for local review.

Even time-sensitive volcanic or landslide warnings in Ecuador move through leaders, assemblies, radio, and phones because community members may be more likely to act on information from trusted leaders than unfamiliar sources. AI could support rather than replace these channels by reading satellite imagery for early landslide signs, translating warnings rapidly into Kichwa, and routing them first to local leaders, who determine how to communicate the threat. The goal is not a universal technical model but tools that move at the "speed of trust": a pace set by trusted channels checking information before wider circulation.

Governance also means moving beyond single-source data collection. AI can accept voice notes, maps, and photographs rather than only fixed survey forms. These formats already exist; what remains is the choice to treat them as valid input.

What's at stake

These examples show that democratizing AI is not simply a matter of access or connectivity. Systems serving these communities must increasingly process mixed-language speech, voice, images, and other inputs, but technical capability does not ensure sociocultural fit or legitimate governance. A system may recognize words, symbols, or seasonal references without knowing when knowledge should be used, who may interpret it, or whether it should be shared. The question is therefore whether AI will continue imposing one "universal" model of knowledge or can be designed and governed across different ways of knowing without flattening the social, territorial, and collective relations that give them meaning.

A better path is possible. WhatsApp-based assistants are promising not because they are new but because they use a platform communities already know. Designed for group use, they could support crop planning, biodiversity monitoring, language revitalization, and community monitoring of land use, environmental threats, and changes in shared natural resources. Yet familiarity with the platform is insufficient: communities still need agreed rules defining what data are collected, who can access them, and how resulting information may be shared or acted upon.

This is possible now if designers and policymakers build around the lives and knowledge of the communities they seek to serve. What is at stake is more than access: it is the right to shape how knowledge is made, used, and acted upon in an algorithmic world. Ignoring Indigenous worldviews would reproduce institutional mistakes of the past; genuine partnership could instead allow AI to earn something few technologies achieve, not only adoption but also trust.

ORCID iDs

Inti Josue Cachipuendo-Contero https://orcid.org/0009-0009-4557-4300

Justin Yeager https://orcid.org/0000-0001-8692-6311

Author contributions

Authors contributed equally to this work.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by Universidad de Las Américas (UDLA) (grant no. 483.A.XIV.24; JY).

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data availability statement

All data are available in the main text.

References

Albahri AS, Khaleel YL, Habeeb MA, et al. (2024) A systematic review of trustworthy artificial intelligence applications in natural disasters. *Computers and Electrical Engineering* 118: 109409.

Andrade Ciudad L and Howard R (2021) Las lenguas quechuas en tres países andino-amazónicos: De las cifras a la acción ciudadana. *Káñina* 45(1): 7–38.

Cerrón-Palomino R (1987) *Lingüística Quechua*. Centro de Estudios Rurales Andinos "Bartolomé de Las Casas".

Chávez Velásquez CR, Sinaluisa Pilco AM, Lema Palaquibay LF, et al. (2024) The heritagescape of Kichwa people of Nizag built upon traditional plant usage along a Chimborazo variant of the Andean Road System or Qhapaq Ñan. *Geographies* 4: 537–562.

Gqweta N (2025) Psychotherapy in South African indigenous languages: Positioning isiXhosa as a language of symbolisation. *South African Journal of African Languages* 45(3): 354–367.

Instituto Nacional de Estadística y Censos (INEC) (2023) *Resultados del VIII Censo de Población y VII de Vivienda 2022*. INEC. Available at: https://www.censoecuador.gob.ec (accessed 20 August 2026).

Katz R, Valencia R, Callorda F, et al. (2024) *Impacto de la conectividad digital en hogares liderados por mujeres, individuos de pueblos indígenas o afrodescendientes en Ecuador*. Inter-American Development Bank. Available at: https://doi.org/10.18235/0012873.

Lewis JE, Whaanga H and Yolgörmez C (2025) Abundant intelligences: Placing AI within Indigenous knowledge frameworks. *AI & Society* 40: 2141–2157.

Perera M, Vidanaarachchi R, Chandrashekeran S, et al. (2025) Indigenous peoples and artificial intelligence: A systematic review and future directions. *Big Data & Society* 12(2): 1–22.

Requelme N, Cachipuendo C, Alba G, et al. (2024) Pueblo Kayambi: Planes de vida comunitarios, derecho a la tierra y al territorio. In: Guarderas AB, Espín MA, Granda Merchán S et al. (eds) *Diálogos para la paz: miradas desde la diversidad*. Abya-Yala, 575–601. Available at: https://doi.org/10.17163/abyaups.80.595.

Taguchi C, Saransig J, Velásquez D, et al. (2024) Killkan: The automatic speech recognition dataset for Kichwa with morphosyntactic information. In: Proceedings of the 2024 Joint International *Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Torino, Italia: ELRA and ICCL, pp. 9753–9763. Available at: https://aclanthology.org/2024.lrec-main.852/.

Usmanova S, Palaniappan A, Sathiyamurthy A, et al. (2025) Cross-lingual information processing in Indigenous language archives. *Indian Journal of Information Sources and Services* 15(4): 292–301.