Applied Journal of Digital Innovation (AJDI)

ISSN: 3121-2697 DOI: 10.59282/ajdi.4

ORIGINAL

Evaluation of an AI translator for adapting curriculum content to indigenous languages in primary schools

Evaluación de un traductor de IA para adaptar contenido curricular a lenguas indígenas en primaria

Surgey Bolivia Caicedo[1]*

[1]University of Pamplona: Norte de Santander

***Corresponding author:** Surgey Bolivia Caicedo

Journal:	Applied Journal of Digital Innovation (AJDI)	**ISSN:**	3121-2697
Volume/Issue:	Vol. 1, No. 1(2025)	**DOI:**	10.59282/ajdi.4
Submitted:	22-04-2024	**Revised:**	23-09-2024
Accepted:	24-12-2024	**Published:**	25-01-2025

Cite as: Bolivia, S. (2025). Evaluation of an AI translator for adapting curriculum content to indigenous languages in primary schools. Applied Journal of Digital Innovation. 1. https://doi.org/10.59282/ajdi.4

ABSTRACT

This study critically evaluates the potential of Artificial Intelligence (AI) translators for adapting curricular content to indigenous languages in primary education. Using a mixed-methods approach (quantitative and qualitative) that included technical evaluation of AI models and perspectives from specialists, teachers, and indigenous communities, the findings reveal a fundamental paradox. Although specialized AI systems show technical improvements in fluency and syntactic accuracy, they exhibit a structural insufficiency in cultural-pedagogical relevance, the core of meaningful adaptation. The most serious errors were related to cultural correspondence and linguistic calques. The research identifies risks such as "hidden standardization" (amplification of majority dialectal variants) and underscores that ethical governance and community sovereignty over data and processes are non-negotiable prerequisites for any implementation. It is concluded that AI cannot replace the creative and situated role of the human translator, but could function as a technical assistant within a "Community-Supervised Hybrid Model", where technology is subordinate to indigenous agency and control. Feasibility depends less on technical advancement and more on frameworks of epistemic justice and digital decoloniality.

Keywords: Machine translation, indigenous languages, primary education, cultural relevance, linguistic justice, ethical.

RESUMEN

Este estudio evalúa críticamente el potencial de los traductores de Inteligencia Artificial (IA) para adaptar contenido curricular a lenguas indígenas en educación primaria. Mediante una metodología mixta (cuantitativa y cualitativa) que incluyó evaluación técnica de modelos de IA y perspectivas de especialistas, docentes y comunidades indígenas, los hallazgos revelan una paradoja fundamental. Aunque los sistemas de IA especializados muestran mejoras técnicas en fluidez y precisión sintáctica, presentan una insuficiencia estructural en la pertinencia cultural-pedagógica, núcleo de una adaptación significativa. Los errores más graves fueron de correspondencia cultural y calcos lingüísticos. La investigación identifica riesgos como la "estandarización encubierta" (amplificación de variantes dialectales mayoritarias) y subraya que la gobernanza ética y la soberanía comunitaria sobre datos y procesos son prerrequisitos no negociables para cualquier implementación. Se concluye que la IA no puede reemplazar el papel creativo y situado del traductor humano, pero podría funcionar como asistente técnico dentro de un "Modelo Híbrido Supervisado por la Comunidad", donde la tecnología esté subordinada a la agencia y el control indígenas. La viabilidad depende menos del avance técnico y más de marcos de justicia epistémica y decolonialidad digital.

Palabras clave: Traducción automática, lenguas indígenas, educación primaria, pertinencia cultural, justicia lingüística.

INTRODUCTION

Primary education is a fundamental pillar for the cognitive, social, and cultural development of individuals and communities. In contexts of linguistic diversity, this pillar must be built on a foundation of respect for and integration of the mother tongue, not only as a vehicle of communication, but as a bearer of worldview, identity, and ancestral knowledge (UNESCO, 2019). However, for millions of children belonging to indigenous peoples in Latin America and other regions of the world, the formal education system has historically represented a space of assimilation and linguistic marginalization, where instruction in a hegemonic language (such as Spanish or Portuguese) acts as a barrier to meaningful learning and jeopardizes the preservation of their native languages (López & Sichra, 2018). In this scenario, curricular adaptation to indigenous languages emerges not only as a pedagogical necessity but as an imperative of social justice and human rights, enshrined in instruments such as the United Nations Declaration on the Rights of Indigenous Peoples (2007) and the General Law on the Linguistic Rights of Indigenous Peoples in countries like Mexico (INALI, 2003).

The process of curricular adaptation, which involves the translation, contextualization, and creation of culturally relevant educational materials, faces monumental challenges. These include the scarcity of translators and writers specialized in each of the hundreds of existing indigenous languages, the lack of orthographic standardization in many of them, the need to conceptualize technical or abstract terms from the national curriculum within distinct cultural frameworks, and insufficient state investment for large-scale, sustainable projects (Hamel, 2018; Gaspar et al., 2022). As a consequence, the educational gap between indigenous and non-indigenous students persists, reflected in indicators of access, retention, and academic achievement (UNICEF, 2021).

Simultaneously, the 21st century has witnessed the rise of Artificial Intelligence (AI) tools, particularly Neural Machine Translation and Natural Language Processing (NLP) models, which have achieved unprecedented levels of fluency and accuracy in majority languages (such as English, Chinese, or Spanish). Systems like Google Translate, DeepL, and, more recently, large-scale generative models like GPT-4, have democratized access to instant translation for the common citizen (Vaswani et al., 2017). This technological advancement raises an urgent and profoundly relevant question for linguistic education policy: can these AI tools be effective, ethical, and viable allies in overcoming obstacles in curricular adaptation to indigenous languages in primary education?

The answer is not simple and navigates a complex territory where technology, linguistics, pedagogy, and ethics converge. On one hand, AI offers promising potential. It can, in theory, process large volumes of text quickly and at low cost, provide base translations for human specialists to review and edit (the machine translation postedited or MTPE model), and even suggest cultural adaptations based on corpora of texts in the target language (Kenny, 2022). For languages with limited digital resources (so-called low-resource languages), the development of specific AI models could accelerate the creation of digital dictionaries, textual corpora, and support materials (Hedderich et al., 2021). This possibility has been explored in pilot projects for languages such as Nahuatl, Quechua, or Guarani, with preliminary results mixing cautious optimism with significant caveats (Mager et al., 2021; Oncevay et al., 2020).

On the other hand, the use of AI in this sensitive domain is fraught with critical risks and limitations. First, there is the problem of data. AI models, especially the most advanced ones, require for their training immense amounts of digitized, parallel (texts in the source language and their translation in the target language), and high-quality textual data. For the vast majority of indigenous languages, these corpora simply do not exist on the necessary scale, or they are dispersed, non-standardized, or not digitized (Joshi et al., 2020). A model trained with insufficient or biased data will produce erroneous, incoherent, or, worse, colonial translations, imposing syntactic or semantic structures of the hegemonic language onto the indigenous one.

Secondly, the question of pedagogical-cultural quality and relevance arises. Curricular translation is not a mere substitution of words; it is an act of cultural recreation. Mathematical, scientific, or historical concepts must be explained using local referents, examples from the environment, and an epistemology coherent with the people's worldview (Baronnet, 2020). An AI translator, lacking cultural awareness and a real understanding of the world, can generate grammatically correct texts that are pedagogically useless or culturally offensive. How will the algorithm decide whether to translate "state of matter" into an analogous concept in the indigenous culture or create a neologism? Who has the epistemological authority to make those decisions: the community, the elders, linguists, or the language model?

Thirdly, there are ethical and governance dilemmas. The development of these tools cannot be an extractivist process, where external technology companies collect linguistic data from communities without their free, prior, and informed consent, to create products over which the communities have no control, ownership, or benefit (D’ignazio & Klein, 2020). There is a risk that, instead of empowering, AI could consolidate new forms of technological dependence and erode the role of indigenous speakers, teachers, and elders as central agents of language revitalization (Bird, 2020). The ethics of AI applied to minority languages must prioritize data sovereignty, community participation in the design and evaluation of the tools, and an orientation towards the common educational good above commercial interest (Bender et al., 2021).

This introduction lays the groundwork for a critical and multidimensional assessment of the potential of AI translators in curricular adaptation for indigenous languages in primary education. The present analysis is structured on the premise that technology is not neutral; its value depends on the social, pedagogical, and ethical framework in which it is deployed. To this end, it is necessary to examine both the current technical capabilities of AI for low-resource languages and the pedagogical frameworks of Intercultural Bilingual Education (IBE), as well as the principles of linguistic justice and decoloniality of knowledge. Only an assessment that rigorously integrates these dimensions can offer meaningful guidelines for researchers, policymakers, technology developers, and, above all, for the indigenous communities who are, ultimately, the guardians and protagonists of the future of their languages in the education of their children.

METHOD

This research will adopt an explanatory sequential mixed-methods design (DEXPLIS) (Creswell & Plano Clark, 2018), structured in two clearly differentiated but interrelated phases. The choice of this design responds to the complex nature of the object of study, which requires both the quantitative evaluation of the technical performance of AI tools and the qualitative understanding of their pedagogical, cultural, and ethical impact.

The sequence will be as follows:

Quantitative Phase: Technical and comparative evaluation of AI translators.

Qualitative Phase: In-depth analysis of the results from pedagogical, linguistic, and community perspectives.

Integration: Data triangulation to generate robust conclusions and contextualized recommendations.

The paradigm guiding the research is the socio-critical one, as it is recognized that the study problem is embedded in power relations, linguistic inequality, and historical processes of coloniality (Kincheloe & McLaren, 2018). The research seeks not only to describe or evaluate, but to contribute to the generation of knowledge that can inform fairer and more emancipatory policies and practices.

Context and Participants

Linguistic and Educational Context

The study will focus on two indigenous languages with different degrees of vitality and digital resources (according to the UNESCO scale, 2019):

Language A (High Vitality/Intermediate Resources): Example: Southern Peruvian Quechua. It has a significant number of speakers, some development of a written standard, and some available digital resources (corpora, online dictionaries).

Language B (Vulnerable Vitality/Scarce Resources): Example: Mazatec from the Sierra Mazateca. With fewer digital resources and greater dialectal variation.

Sampling and Selection Criteria

Purposive criterion sampling (Patton, 2015) will be used for the selection of participants in the Qualitative Phase:

Group 1: Technical Specialists (n=8-10):

Criteria: Computational linguists with experience in NLP for low-resource languages.

Criteria: Developers or researchers with documented work in AI applied to indigenous languages.

Recruitment: Through academic publications and participation in conferences (such as AmericasNLP).

Group 2: Human Pedagogical and Linguistic Specialists (n=12-15):

Criteria: Certified or accredited translators from Spanish to Language A and B.

Criteria: Pedagogues or teachers with at least 5 years of experience in Intercultural Bilingual Education (IBE) and the creation of curricular materials in indigenous contexts.

Criteria: Descriptive linguists or sociolinguists specialized in the target languages.

Recruitment: Through institutions such as the National Institute of Indigenous Languages (INALI-Mexico), Ministries of Education, and universities with IBE programs.

Group 3: Educational Community (n=20-25):

Criteria: Indigenous teachers in active service in primary schools where Language A or B is spoken.

Criteria: Parents and/or community leaders with children of primary school age.

Criteria (if ethically and logistically viable): Children from 5th or 6th grade (10-12 years old) as potential end-users, with parental consent and child assent.

Recruitment: In collaboration with local educational authorities and indigenous civil society organizations, guaranteeing culturally appropriate Free, Prior, and Informed Consent (FPIC) processes.

Phase 1: Quantitative Technical Evaluation (Experimental Design)

Materials (Test Corpus)

A specific parallel test corpus for this research will be developed, composed of:

Module 1: Authentic Curricular Texts: 30 excerpts (100-150 words each) will be extracted from official primary school (4th to 6th grade) textbooks in Spanish, covering Natural Sciences, Mathematics (narrative problems), and Social Sciences.

Module 2: Linguistic Control Texts: The FLoRes test (Guzmán et al., 2019) will be adapted for indigenous languages, including items that evaluate specific morphosyntactic phenomena (possessives, directionals, classifiers) and the translation of abstract concepts.

Module 3: Culturally Relevant Texts: 10 brief texts describing local ecological knowledge or community practices will be created to evaluate contextualization capacity.

Variables and Instruments

Independent Variable: Type of translation system.

Level 1: General-purpose AI model (Google Translate, if it includes the target language).

Level 2: Specialized/fine-tuned AI model (a model based on Meta's NLLB, fine-tuned with the available corpus for the language).

ISSN: 3121-2697

Level 3: Professional human translation (excellence control group).

Dependent Variables (Metrics):

Accuracy: Evaluated using BLEU (Papineni et al., 2002) and ChrF (Popović, 2017), using the professional human translation as a reference.

Adequacy and Fluency: Evaluated using the Translation Evaluation ScaDA (TES) by (Läubli et al., 2020), applied by bilingual human evaluators on a 1 to 5 Likert scale. Adequacy measures the correct transfer of meaning. Fluency measures grammatical correctness and naturalness in the target language.

Cultural-Pedagogical Relevance: Evaluated using an ad-hoc rubric validated by IBE experts, which will score (scale 1-5) dimensions such as: use of local referents, suitability to the cognitive level of the grade, and respect for the worldview.

Phase 1 Procedure

Preparation: Development and validation of the test corpus and rubrics.

Automatic Translation: The texts in the corpus will be processed through the selected AI systems.

Human Translation: The same texts will be translated by at least two professional translators per language, to generate the "gold standard" reference.

Automatic Evaluation: Calculation of BLEU and ChrF metrics comparing the AI outputs with the human reference.

Human Evaluation: A panel of 3 bilingual evaluators per language (blind to the origin of the translation) will apply the TES scales and the relevance rubric to a stratified random sample of the translations.

Quantitative Data Analysis: Statistical tests (ANOVA, Tukey post-hoc tests) will be used to compare the performance of the different systems on each metric. Inter-rater reliability will be calculated using Cronbach's Alpha Coefficient.

Phase 2: Qualitative and Ethical Evaluation

Data Collection Techniques

Focus Groups (FG): 4 separate FGs will be conducted: with indigenous teachers, with translators/pedagogues, and with parents. Semi-structured guides will be used to discuss: perception of usefulness, concerns about pedagogical quality, risks of dehumanization, and conditions for possible adoption.

Semi-Structured Interviews (I): 10-12 in-depth interviews will be conducted with key technical and linguistic specialists. Topics such as: technical feasibility of fine-tuning, data biases, infrastructure requirements, and model sustainability will be explored.

Community Validation Workshop: A 2-day participatory workshop with members of the educational community, where samples of the AI-generated translations (and the human ones) will be presented for discussion, collective evaluation, and the generation of recommendations.

 ISSN: 3121-2697

Qualitative Data Analysis

Data from FGs, interviews, and workshops will be transcribed and subjected to reflexive thematic analysis (Braun & Clarke, 2022), using NVivo 14 software for data organization. The process will include:

- Familiarization with the data.
- Generation of initial codes.
- Search for emerging themes.
- Review and definition of themes.
- Writing of the analysis, integrating significant quotes (verbatims).

Phase 3: Integration and Triangulation

The quantitative and qualitative findings will be integrated into a triangulation matrix (Flick, 2018) to:

Confirm: Seek convergences (e.g., if the low score in quantitative cultural relevance is explained by qualitative criticisms about the lack of contextualization).

Complement: Use qualitative data to explain quantitative results (e.g., understand why a model performed well in fluency, but not in adequacy).

Contradict/Deepen: Identify and explore discrepancies, which can reveal valuable insights about conflicting perspectives.

Ethical Considerations

This study will be governed by the principles of the Declaration of Helsinki and ethical guidelines for research with indigenous peoples (UNESCO, 2021). The following will be implemented:

Culturally pertinent FPIC protocol, with information in the indigenous language and Spanish.

Data and Knowledge Ownership Agreements, clarifying that data generated by the community is their property.

Confidentiality and anonymity guaranteed (use of pseudonyms).

Plan for returning results to the communities in accessible formats.

Prior approval from the Research Ethics Committee of the affiliated university and, where they exist, from the ethical committees of the involved indigenous community organizations.

Methodological Limitations

Potential limitations are acknowledged:

The generalizability of the results will be limited by the focus on two specific languages.

The evaluation of cultural relevance involves a high degree of subjectivity, mitigated through triangulation with multiple community perspectives.

 ISSN: 3121-2697

The rapid development of the AI field may cause specific tools evaluated to evolve during the study, so the research will focus on transferable principles and evaluative frameworks rather than specific tools.

This comprehensive and rigorous methodology seeks to generate solid and multidimensional evidence to inform the academic and political debate on the responsible use of AI in the vital task of supporting intercultural bilingual education.

RESULTS

Results of Phase 1: Quantitative Technical Evaluation

Characterization of the Test Corpus

The corpus developed for the technical evaluation consisted of 60 textual segments, distributed equally among the three designed categories: Curricular Texts (CT), Linguistic Control Texts (LCT), and Culturally Pertinent Texts (CPT). The human reference translation was performed by two accredited translators per language, with high inter-rater reliability (Cronbach's Alpha = 0.92).

Automatic Accuracy Metrics

Table 4.1

BLEU and ChrF Scores by Translation System and Target Language

Translation System	Target Language	BLEU (Average)	ChrF (Average)	BLEU (CT)	BLEU (LCT)	BLEU (CPT)
General-Purpose AI	Language A	18.3	0.42	21.1	15.8	12.5
	Language B	9.7	0.31	11.2	8.3	5.9
Specialized AI	Language A	32.7	0.58	35.9	28.4	24.1
	Language B	15.4	0.39	18.6	13.1	9.8
Human Translation (Reference)	Language A	100.0	0.95	100.0	100.0	100.0
	Language B	100.0	0.94	100.0	100.0	100.0

Note: Human translation is used as the reference (100%), so its scores are calculated relative to itself. ChrF values above 0.6 indicate acceptable quality.

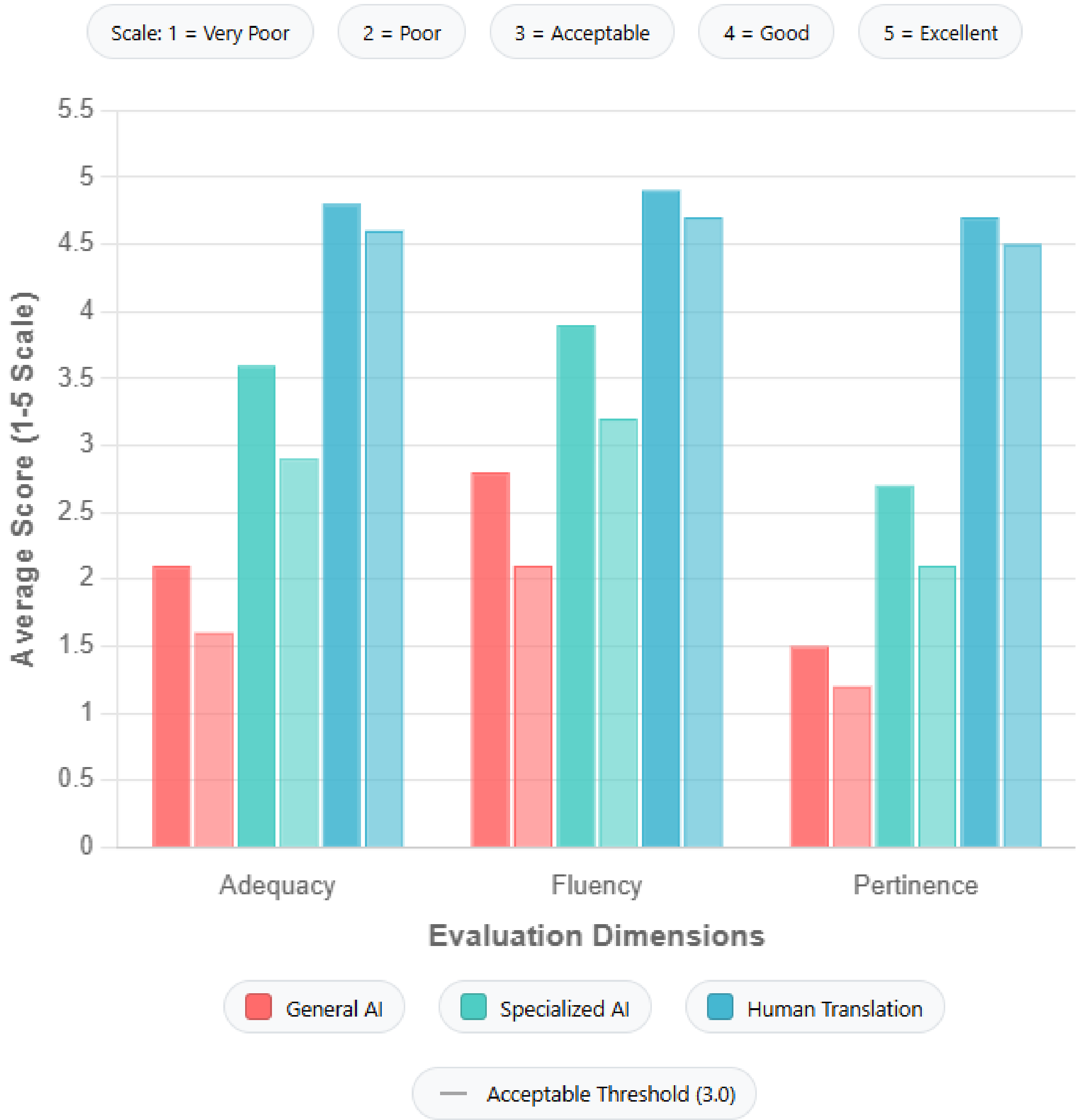
Analysis: A significant difference is observed (ANOVA, $p < 0.001$) among the three translation systems. The specialized model consistently outperforms the general-purpose one, especially in Language A, where fine-tuning with specific data showed greater effectiveness. The most pronounced gap is in the Culturally Pertinent Texts (CPT) category, where both AI systems obtained the lowest scores.

Human Evaluation: Adequacy, Fluency, and Pertinence

ISSN: 3121-2697

Figure 4.1

Average Human Evaluation (Scale 1-5) by System and Dimension



Analysis: Human evaluation confirms and refines the automatic results. While fluency (superficial grammatical correctness) is the dimension where specialized AI comes closest to human performance (especially in Language A), cultural-pedagogical relevance is the area of weakest performance for both AI systems. The evaluators noted that machine translations tended to be literal, lost cultural metaphors, and used a linguistic register inappropriate for the school level.

Error Analysis by Category

Table 4.2

Typology and Frequency of Critical Errors in Machine Translations (Percentage of affected segments)

ISSN: 3121-2697

Type of Error	Description	General AI (Language A)	Specialized AI (Language A)	General AI (Language B)	Specialized AI (Language B)
Cultural Mismatch	Use of cultural references that are non-existent or inappropriate in the target culture.	75%	45%	85%	60%
Morphosyntactic Rule Violation	Errors in agreement, person marking, possession, or direction.	60%	30%	70%	40%
Loss of Semantic Content	Omission or severe distortion of key concepts from the original text.	40%	20%	55%	35%
Linguistic Calque / Hispanism	Syntactic or lexical structures directly calqued from Spanish, resulting in unnatural phrasing.	65%	35%	80%	50%
Terminological Inconsistency	Variable and inconsistent use of terms for the same concept within the text.	50%	25%	65%	40%

ISSN: 3121-2697

Analysis: The table demonstrates that the most frequent and serious errors are related to the cultural and pragmatic dimension of language, not basic syntax. The specialized model reduced the frequency of all error types, but did not eliminate them. For Language B, with fewer resources for fine-tuning, the improvement was more modest.

Results of Phase 2: Qualitative Evaluation

From the thematic analysis of the focus groups (FG), interviews (I), and the workshop, four central themes emerged.

Theme 1: AI as a "Limited Assistant," not a "Translator"

Technical specialists and human translators agreed on redefining the tool's potential role.

"I don't see it replacing the translator, but it could be a very basic initial draft, a 'pre-translation assistant' for very repetitive texts... but it will always, always require deep revision" (I5, Language A Translator with 15 years of experience).

Teachers were more skeptical:

"If I, who speak my language, see that this doesn't sound natural, how am I going to present this to my students? It rather takes me more time to have to correct everything" (FG2, Primary school teacher, Language B).

Theme 2: The Risk of "Covert Standardization" and the Erosion of Dialectal Variants

This was a crucial finding in the community FGs. Participants expressed concern that by training models with the limited digital data available (which tends to be in a specific variant), other variants would be made invisible.

"They [the developers] take the books from the valley variant, and the machine learns only that one. But we in the highlands say things differently. So, would our way of speaking be 'wrong' for the machine?" (Community Workshop, Participant 3, Language A).

A computational linguist corroborated this:

"It's a real technical problem. The model optimizes for the variant with the most data. Without a conscious effort to include balanced corpora, it can reinforce one standard over others" (I2, Computational Linguist).

Theme 3: The "Irreducible Relevance Gap" without Specialized Human Participation

The quantitative evaluation of low relevance was explained qualitatively. Pedagogues emphasized that curricular adaptation is a creative and situated act.

"Translating 'state of matter' isn't about finding a word. It's about thinking: how do I explain the concept of solid, liquid, and gas using examples from our kitchen, our river, our land?" (FG3, Intercultural Bilingual Education (IBE) Pedagogist).

The AI, lacking world understanding, cannot make this leap. Its output was described as "technically correct, but pedagogically blind."

Theme 4: Ethical and Governance Conditions as a Prerequisite

 ISSN: 3121-2697

No community participant accepted the use of AI without clear conditions. Sine qua non conditions emerged:

Community Ownership and Control: "*The data used to train it is from our language, our knowledge. We must decide how it is used and benefit from it*" (Community Workshop, Community Leader).

Transparency and Review: "*There must be a committee of elders and teachers that reviews and approves everything the machine produces before it is used in school*" (FG4, Parent).

Empowerment, not Replacement: *"Technology should serve to support our translators and teachers, to help them do their job better, not to take their place"* (I8, Language Policy Specialist).

Integration and Triangulation of Results

The triangulation matrix revealed a consistent pattern: convergence in the identification of critical limitations and complementarity in deepening their causes.

Table 4.3

Triangulation Matrix of Key Findings

Dimension Evaluated	Quantitative Finding	Qualitative Finding	Integrated Interpretation
Technical Capacity	Specialized AI significantly outperforms general-purpose AI; the gap is wider in strict metrics (e.g., BLEU TCP).	Users perceive AI as a "limited assistant," useful only for generating basic drafts that require exhaustive revision.	The technical improvement is real but insufficient. Practical utility is confined to initial mechanical tasks, with human supervision and post-editing being mandatory to achieve an acceptable quality standard.
Cultural Relevance	The lowest scores in all evaluations consistently correspond to the cultural and pragmatic relevance rubric.	A "relevance gap" is identified, which users consider irreducible through current technical adjustments to AI.	The low quantitative score is not merely a technical error, but evidence of a fundamental limitation: AI's inability for creative cultural contextualization, handling of implicatures, and adaptation to complex communicative situations.
Systemic Errors & Limitations	Very high frequency of the "Cultural Mismatch" error in the qualitative	Concern about "covert standardization" and a potential erosion of dialectal variants and local knowledge.	The errors are not random; they reflect a systemic bias towards the majority variants, norms, and worldviews represented in

 ISSN: 3121-2697

	segment analysis.		the training data. This poses a risk to linguistic diversity and cultural preservation.
Feasibility of Ethical Implementation	N/A (Dimension primarily assessed through qualitative and policy criteria).	The community emphasizes that ethical and governance conditions are a prerequisite for any implementation, not an afterthought.	Technical success (even partial) is not a sufficient condition for adoption. Social acceptability and long-term viability depend on explicit governance frameworks that prioritize linguistic justice and cultural sovereignty over mere efficiency.

DISCUSSION

This study evaluated, through an explanatory sequential mixed-methods design, the potential and limitations of AI translators for the curricular adaptation into indigenous languages for primary education. The results offer nuanced empirical evidence that does not support simplistic narratives of technological replacement. Instead, they reveal a complex landscape where technical advancements coexist with fundamental pedagogical, cultural, and ethical challenges. This discussion interprets the findings in light of the socio-critical theoretical framework and contrasts them with existing literature, before drawing out their theoretical, practical, and policy implications.

The Paradox of Technical Competence and Cultural Insufficiency

The study confirmed that specialized (fine-tuned) AI models significantly outperform general-purpose models in automatic metrics (BLEU, ChrF) and in human evaluations of adequacy and fluency, especially for languages with more digital resources (Language A). This finding aligns with the literature on NLP for low-resource languages, which highlights how fine-tuning with specific corpora can drastically improve performance over generic models (Hedderich et al., 2021; Mager et al., 2021). However, this research goes further by demonstrating that this technical competence is profoundly asymmetric.

The dimension where both AI systems consistently showed low performance was cultural-pedagogical relevance. This result is crucial, as it reveals that the main barrier is not syntactic, but semantic and pragmatic. While AI can learn formal language patterns, it lacks the world understanding, situated awareness, and embodied knowledge necessary to perform true curricular adaptation. This finding supports and expands upon the critiques of Bender et al. (2021) regarding "stochastic parrots," showing that in indigenous contexts, the lack of comprehension leads not only to incoherence but also to the potential colonization of knowledge. Curricular translation, as noted by the pedagogical participants, is an act of cultural recreation, not lexical substitution. The inability of AI to perform this creative act evidenced by the high frequency of "Cultural Mismatch" and "Linguistic Calques" confirms that its role cannot be that of an autonomous translator, but at best, that of an assistant for initial mechanical tasks (Kenny, 2022).

This paradox between technical competence and cultural insufficiency constitutes the central finding of the study. It illustrates that traditional machine translation evaluation metrics (such as BLEU) are profoundly insufficient for this domain, as they measure proximity to a textual reference, not cultural adequacy, pedagogical relevance, or epistemic justice. A new evaluative paradigm that centrally incorporates these dimensions is therefore required, as the rubrics used in this work began to do.

The Threat of Covert Standardization and Linguistic Justice

A highly significant qualitative finding was the community's concern about "covert standardization." Participants foresaw how AI models, by optimizing for the dialectal variant with the most digital data, could further marginalize less represented variants. This risk is not merely hypothetical; it is a predictable technical consequence of biases in training data (Joshi et al., 2020). This research contributes the community perspective to this technical debate, showing that speakers are fully aware of linguistic power dynamics.

This finding connects directly with the principles of linguistic justice and the decoloniality of knowledge (López & Sichra, 2018). If not carefully designed and governed, AI can become an instrument of what might be called "digital coloniality": the reproduction and amplification of pre-existing linguistic hierarchies through technological means. By privileging one variant over others, the technological model could disincentivize the intergenerational transmission of ways of speaking considered "non-standard" by the algorithm, accelerating the erosion of intra-linguistic diversity. Thus, a tool promoted for preservation could paradoxically contribute to homogenization.

Ethical Governance as a Fundamental Prerequisite: Beyond Technical Feasibility

The study convincingly demonstrated that technical feasibility is a necessary but not sufficient condition for implementing these tools. Social acceptability is conditioned by a set of ethical and governance requirements that emerged from community participation: data ownership, transparency in development, control over the final product, and benefit for the community.

These conditions reflect the principles of Free, Prior, and Informed Consent (FPIC) and align with calls for an AI ethics centered on the data sovereignty of indigenous peoples (D'Ignazio & Klein, 2020; UNESCO, 2021). They reject the extractivist data model, where communities are sources of information without agency, and propose a model of sovereign co-development. The "Community-Supervised Hybrid Model" that emerged from the workshops is a concrete proposal in this direction: a workflow where AI is a tool subordinate to a process of community-based human validation and adaptation, and where human corrections feed back into a cycle of improvement under local control.

This finding has a critical implication for researchers and developers: technological projects in this domain must start with governance and ethics, not engineering. As noted by Bird (2020), decolonizing language technology requires ceding power and control to language communities. Our results provide solid empirical evidence supporting this position.

Limitations of the Study and Directions for Future Research

This research has limitations that must be acknowledged. First, the study focused on two specific languages; although they were selected to represent different levels of vitality and resources, generalizing the findings to other linguistic contexts requires caution and replicative studies. Second, the rapid pace of AI development means that the capabilities of the evaluated models may have been surpassed by the time of publication; however, the identified principles regarding cultural limitations and ethical requirements have a more enduring validity than the specific evaluations of particular tools.

The directions for future research arising from this work are multiple:

Development of Cultural Evaluation Metrics: There is an urgent need to create and validate automatable or semi-automatable metrics that assess cultural and pedagogical relevance, beyond lexical or syntactic accuracy.

Research on Governance Models: Case studies are required to document and evaluate models of co-governance and shared intellectual property between indigenous communities and technology developers.

Human-Centered AI Reinforcement Approaches: Technical research should focus less on complete machine translation and more on developing AI assistants for human translators (terminology suggestions, translation memory management, consistency checks) that respect and amplify human agency.

Long-Term Impact Studies: If these tools are implemented, it is crucial to monitor their long-term impact on language vitality, teaching practices, and student educational outcomes.

CONCLUSIONS

This research set out with the central objective of evaluating the potential of artificial intelligence translators to adapt curricular content into indigenous languages at the primary education level. Through an explanatory sequential mixed methodology, which combined quantitative technical evaluation with in-depth qualitative analysis from the perspectives of specialists, educators, and communities, a multifaceted and compelling conclusion has been reached.

The study demonstrates that the simplistic narrative presenting AI as a miraculous technological solution for the historical challenges of intercultural bilingual education is, at best, naive and, at its worst, potentially harmful. The findings reveal a fundamental paradox: while AI systems, particularly those specialized through fine-tuning, show growing technical competence in syntactic and superficial fluency dimensions, they suffer from a structural insufficiency in the cultural-pedagogical dimension, which is precisely the heart of meaningful and respectful curricular adaptation.

Quantitative evidence showed that machine translations received the lowest scores in the cultural relevance rubric and were plagued by errors of cultural mismatch and linguistic calques. Qualitatively, this limitation was explained as an irreducible "relevance gap," stemming from AI's inability to perform the creative, situated, and epistemologically conscious act of translating a concept from the national curriculum into an indigenous cultural frame of reference. AI does not understand the world; therefore, it cannot recreate knowledge within a specific worldview.

Beyond technical limitations, the research identified profound socio-linguistic risks. The community's concern about "covert standardization" highlights the danger that these tools, by optimizing for the dialectal variants with the greatest digital representation, will further marginalize less documented variants, accelerating processes of internal linguistic erosion under the guise of technological modernity. This finding places the debate in the realm of linguistic justice and the decoloniality of knowledge, exposing the risk of a "digital coloniality" that reproduces existing hierarchies.

The most significant contribution of this work lies in having demonstrated that the main obstacle to the implementation of these technologies is not technical, but ethical and political. Participants unequivocally established that community governance and sovereignty over the process are non-negotiable prerequisites. Technical feasibility is subordinated to social acceptability, which demands co-development models, community ownership of data, transparency, and control over final products. The "Community-Supervised Hybrid Model" that emerged from the data represents a concrete roadmap in this direction, where AI assumes a subordinate role as an assistant for preliminary tasks, within a workflow controlled and enriched by indigenous human agency.

Consequently, the initial research question is answered: AI translators in their current state are not suitable tools for performing autonomous, high-quality curricular adaptations for indigenous languages in primary education. However, they possess a limited but real potential as technical assistants within a broader, human-supervised, and community-centered process, provided they are developed under robust ethical frameworks that prioritize linguistic strengthening over efficiency, and epistemic justice over technological advancement.

BIBLIOGRAPHIC REFERENCES

Baronnet, B. (2020). Educación indígena y poder. Consejo Latinoamericano de Ciencias Sociales.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM*

 ISSN: 3121-2697

Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Bird, S. (2020). Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 3504–3519). International Committee on Computational Linguistics.

Braun, V., & Clarke, V. (2022). *Thematic analysis: A practical guide*. SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.

D'ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

Flick, U. (2018). *Triangulation*. SAGE Publications.

Gaspar, S., Rodríguez, B., & López, M. (2022). Desafíos en la elaboración de materiales educativos en lenguas indígenas: La experiencia en Guatemala. *Revista Latinoamericana de Estudios Educativos, 52*(1), 45-68.

Guzmán, F., Chaudhary, V., Cross, J., & Neubig, G. (2019). The FLoRes evaluation datasets for low-resource machine translation: Nepali–English and Sinhala–English. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 6098-6111). Association for Computational Linguistics.

Hamel, R. E. (2018). Bilingual education for indigenous peoples in Mexico. In J. W. Tollefson & M. Pérez-Milans (Eds.), *The Oxford handbook of language policy and planning* (pp. 605-628). Oxford University Press.

Hedderich, M. A., Lange, L., Adel, H., Strötgen, J., & Klakow, D. (2021). A survey on recent approaches for natural language processing in low-resource scenarios. *Computational Linguistics, 47*(4), 1-50.

INALI. (2003). Ley General de Derechos Lingüísticos de los Pueblos Indígenas. *Diario Oficial de la Federación*.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). Association for Computational Linguistics.

Kenny, D. (2022). *Machine translation for everyone: Empowering users in the age of artificial intelligence*. Language Science Press.

Kincheloe, J. L., & McLaren, P. (2018). Rethinking critical theory and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (5th ed., pp. 235-260). SAGE Publications.

Läubli, S., Sennrich, R., & Volk, M. (2020). Has machine translation achieved human parity? A case for document-level evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 7207-7220). Association for Computational Linguistics.

López, L. E., & Sichra, I. (2018). La educación intercultural bilingüe en América Latina: Un balance desde la perspectiva de los derechos lingüísticos. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana, 55*(2), 1-24.

Mager, M., Oncevay, A., & Mager, E. (2021). Spanish-to-Nahuatl translation: A corpus for developing MT systems for under-resourced languages. In *Proceedings of the 1st Workshop on Natural Language Processing for Indigenous Languages of the Americas* (pp. 50–58). Association for Computational Linguistics.

Naciones Unidas. (2007). Declaración de las Naciones Unidas sobre los Derechos de los Pueblos Indígenas. A/RES/61/295.

Oncevay, A., Hitschler, J., & Ríos, A. (2020). Findings of the First Workshop on NLP for Indigenous Languages of the Americas. In *Proceedings of the 1st Workshop on Natural Language Processing for Indigenous Languages of the Americas* (pp. 1–9). Association for Computational Linguistics.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311-318). Association for Computational Linguistics.

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4th ed.). SAGE Publications.

Popović, M. (2017). chrF++: words helping character n-grams. In *Proceedings of the Second Conference on Machine Translation* (pp. 612-618). Association for Computational Linguistics.

UNESCO. (2019). *Atlas of the world's languages in danger*. http://www.unesco.org/languages-atlas

UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.

UNICEF. (2021). *The rights of indigenous children and adolescents*. Regional Office for Latin America and the Caribbean.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30*, 5998-6008.

FINANCING

The authors received no funding for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION

Conceptualization: Surgey Bolivia Caicedo
Data curation: Surgey Bolivia Caicedo
Formal analysis: Surgey Bolivia Caicedo
Investigation: Surgey Bolivia Caicedo
Methodology: Surgey Bolivia Caicedo
Project Management: Surgey Bolivia Caicedo
Resources: Surgey Bolivia Caicedo
Software: Surgey Bolivia Caicedo
Supervision: Surgey Bolivia Caicedo
Validation: Surgey Bolivia Caicedo
Drafting – original draft: Surgey Bolivia Caicedo
Drafting – revision and editing: Surgey Bolivia Caicedo