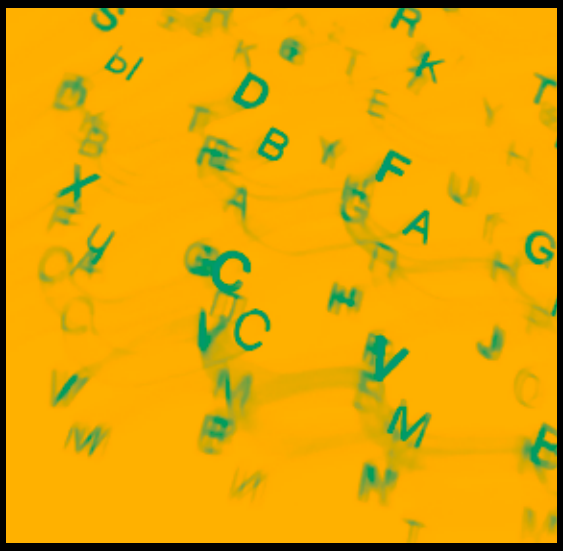
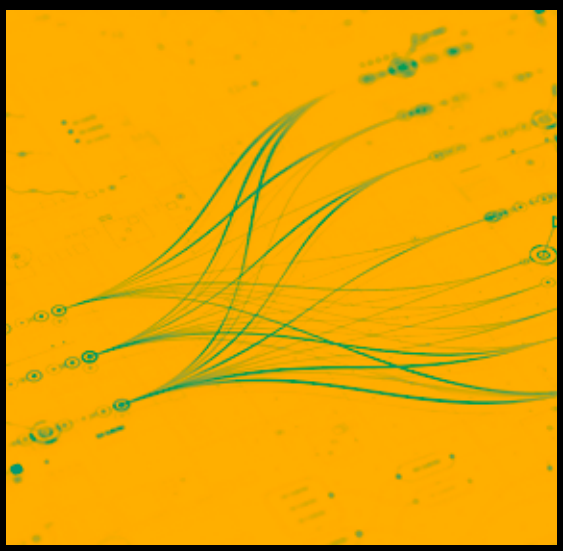
EXECUTIVE SUMMARY

THE PERFORMANCE OF ARTIFICIAL INTELLIGENCE IN THE USE OF INDIGENOUS AMERICAN LANGUAGES





Microsoft IDB | LAB

EXECUTIVE SUMMARY

THE PERFORMANCE OF ARTIFICIAL INTELLIGENCE IN THE USE OF INDIGENOUS AMERICAN LANGUAGES

ASSESSMENT OF THE AI GAP IN INDIGENOUS AMERICAN LANGUAGES

Acknowledgments: We express our sincere gratitude to the interpreters and academics who participated in the project “Native Languages and Their Interaction with Artificial Intelligence.” Their valuable collaboration was essential for evaluating the performance of various artificial intelligence technologies in Indigenous languages, making it possible to highlight the cultural and linguistic gaps present in these tools. In particular, we thank Jacob Cruz, Náhuatl language interpreter; Armando Hueyotenco, Náhuatl language interpreter from the Institute for Transparency, Access to Public Government Information, and Protection of Personal Data of the State of Hidalgo (ITAIH); Ana Paola Quispe Quispe, Aymara language interpreter from the National Academy of the Aymara Language (ANLA); Elmer Machicao, Aymara language interpreter; Tomas Rojas, Mapuche language interpreter; Yony Mediano, Quechua language interpreter; and Mauro Lugo, Guaraní language interpreter. Thanks to their expertise and commitment, significant evaluations were achieved in the Náhuatl, Aymara, Mapuche, Quechua, and Guaraní languages, strengthening the inclusive dialogue between technology and linguistic diversity in the region.

Authors: LLYC Team: Miguel Lucas, Alejandro Burgueño and Miguel Carazas. IDB Team: César Buenadicha, Smeldy Ramírez, César Rosales. Microsoft Team: Daniel Korn, Hiwot Tesfaye, Gretchen Deo.

Design: Alejandro Scaff.

https://bidlab.org

Copyright © 2025 Inter-American Development Bank (“IDB”). This work is subject to a Creative Commons license CC BY 3.0 IGO (https://creativecommons.org/licenses/by/3.0/igo/legalcode). The terms and conditions indicated in the URL link must be met and the respective recognition must be granted to the IDB.

Further to section 8 of the above license, any mediation relating to disputes arising under such license shall be conducted in accordance with the WIPO Mediation Rules. Any dispute related to the use of the works of the IDB that cannot be settled amicably shall be submitted to arbitration pursuant to the United Nations Commission on International Trade Law (UNCITRAL) rules. The use of the IDB’s name for any purpose other than for attribution, and the use of IDB’s logo shall be subject to a separate written license agreement between the IDB and the user and is not authorized as part of this license.

Note that the URL link includes terms and conditions that are an integral part of this license.

The opinions expressed in this work are those of the authors and do not necessarily reflect the views of the IDB, its Board of Directors, of the countries they represent, nor IDB Lab Donors Committee or the countries it represents.

fAIr LAC and the IDB Group's Commitment to Responsible AI

This report was developed within the framework of the *fAIr LAC* initiative through a collaboration between IDB Lab, Microsoft, and LLYC, combining complementary capabilities in research, analysis, and strategy. The publication reflects these organizations; joint commitment to linguistic equity and the responsible use of artificial intelligence in Latin America and the Caribbean.

IDB Lab

IDB Lab is the innovation and venture capital arm of the Inter-American Development Bank Group. We discover new ways to drive social inclusion, environmental action and productivity in Latin America and the Caribbean.
IDB Lab leverages financing, knowledge and connections to support early-stage entrepreneurship, foster new technologies, activate innovative markets and catalyze existing sectors.
www.bidlab.org

AI for Good at Microsoft

The AI for Good Lab at Microsoft is a collaborative effort focused on leveraging artificial intelligence to solve some of the worlds most pressing challenges. By working with global partners and leveraging cutting-edge AI tools, the lab supports innovations in environmental sustainability, agriculture, healthcare, and more. The AI for Good Lab is committed to applying AI solutions to improve lives, drive progress toward the UN's Sustainable Development Goals, and make a lasting impact through technology.

LLYC

LLYC is a global communications, digital marketing, and public affairs consulting firm. Founded in Madrid in 1995 as Llorente y Cuenca, LLYC currently has 20 offices in several countries, including Spain, Argentina, Brazil, and Colombia.
LLYC helps its clients address their strategic challenges with solutions based on creativity, technology, and experience.

fAIr LAC

fAIr LAC is a partnership between the public and private sectors, civil society and academic institutions, designed to influence public policy and the entrepreneurial ecosystem in the promotion of the responsible use of AI.

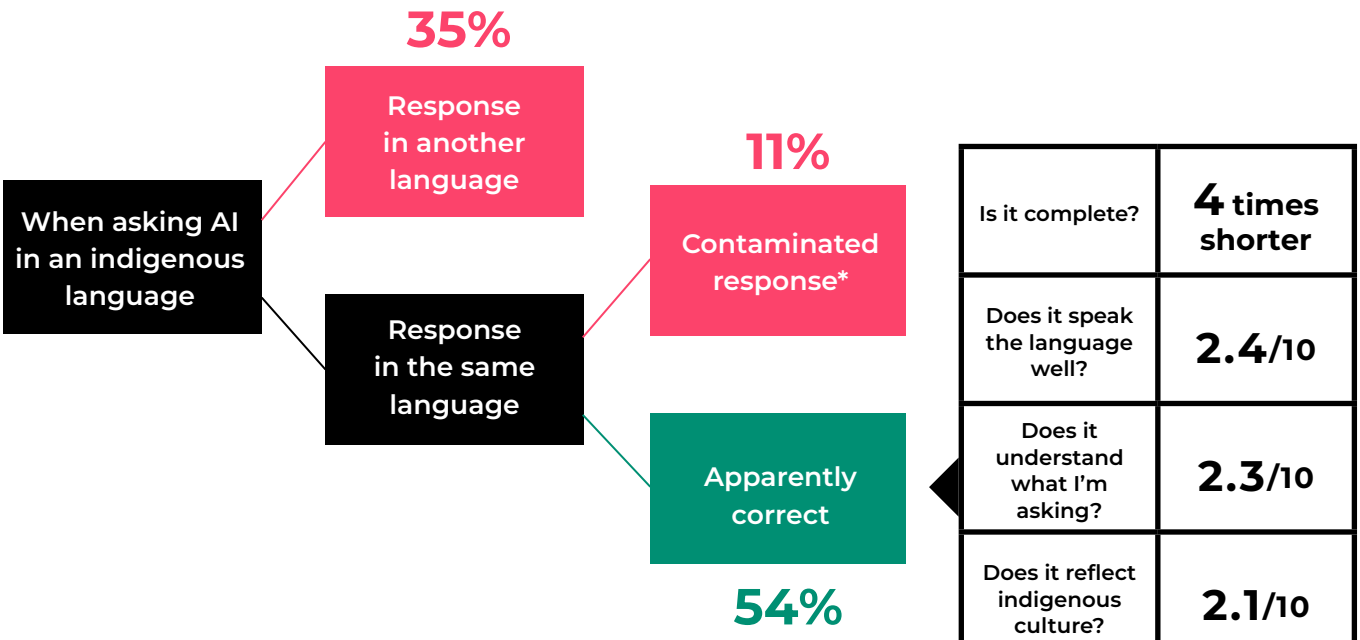
TABLE OF CONTENTS

1. AI performance in indigenous languages 6
2. Impact on community and market 12
3. The enabling environment: key programs and players 15
4. What AI needs to speak a language well 19
5. Technological inclusion strategies 24
6. Recommendations and action plan 28

1. AI PERFORMANCE IN INDIGENOUS LANGUAGES

The journey of a query in an indigenous language to AI

The performance of AI in the use of indigenous languages.



** AI inserts terms from other languages, repeats words in loops, etc.*

In the most favorable scenario (Quechua) AI system fails to meet the passing threshold (3.72); in the least favorable case, (Maya Quiché) model demonstrates near-total incapacity (1.25).

In a language like Catalan, with fewer speakers, AI performs 2.3 times better than in Quechua.

Performance averages across 5 AIs analyzed.

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Spanish	Catalan	Euskera
Idiomatic	4.53	2.68	3.08	3.70	0.42	10.00*	8.62	7.58
Executive	4.48	2.88	2.48	3.77	0.43	10.00*	8.02	6.03
Behavioural	2.15	2.73	2.02	2.80	2.90	10.00*	9.12	7.15
Average	**3.72**	**2.77**	**2.53**	**3.42**	**1.25**	**10.00***	**8.58**	**6.92**

* Spanish shows a score of 10 because it was the reference language for comparisons.

** Chapter 4, page 19 presents the factors that determine the unequal performance of AI for each language.

Proprietary models are twice as effective (x2.2) as open models in indigenous languages

Working to optimize open models in indigenous languages implies a baseline of inferior performance.

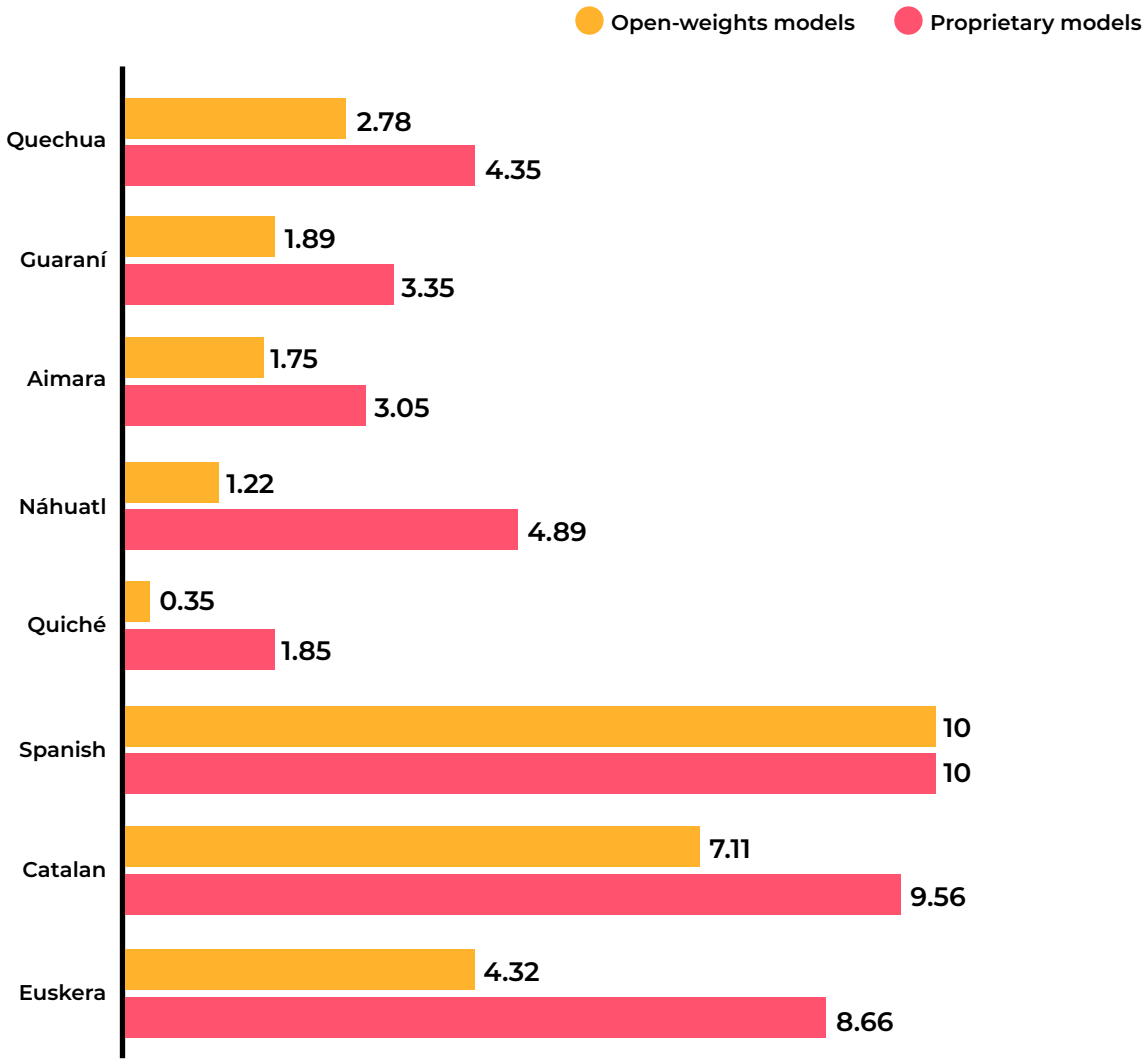
Average performance of proprietary models.

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Spanish	Catalan	Euskera
Idiomatic	5.33	3.28	3.58	5.28	0.67	10.00*	9.83	9.58
Executive	5.36	3.53	3.06	5.39	0.64	10.00*	9.22	7.47
Behavioural	2.36	3.25	2.50	4.00	4.25	10.00*	9.64	8.92
Average	**4.35**	**3.35**	**3.05**	**4.89**	**1.85**	**10.00***	**9.56**	**8.66**

Average performance of open-weights models.

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Spanish	Catalan	Euskera
Idiomatic	3.33	1.79	2.33	1.33	0.04	10.00*	6.79	4.58
Executive	3.17	1.92	1.63	1.33	0.13	10.00*	6.21	3.88
Behavioural	1.83	1.96	1.29	1.00	0.88	10.00*	8.33	4.50
Average	**2.78**	**1.89**	**1.75**	**1.22**	**0.35**	**10.00***	**7.11**	**4.32**

Average performance by AI models.



Expression evaluation: **The fluency of AI expression in indigenous languages is 4 times lower compared to Spanish**

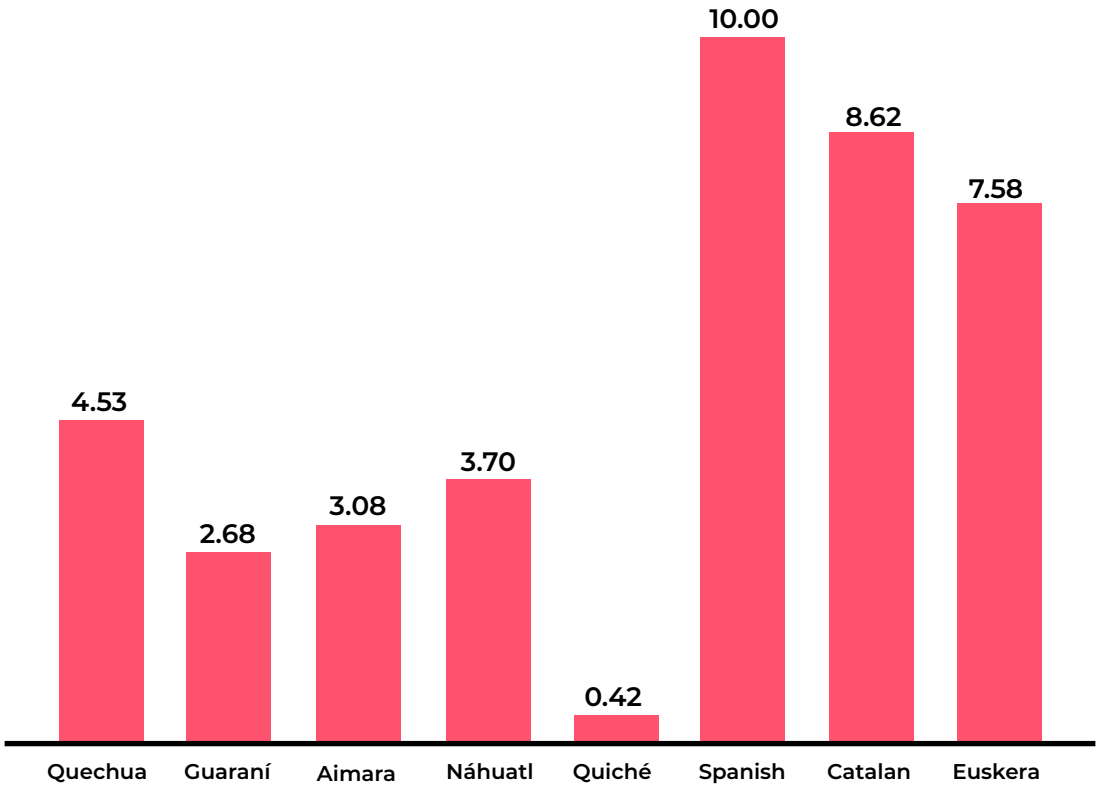
Expression quality degrades as response length increases: coherence is 4.3 times better in Spanish.

Language evaluation. Breakdown of the 4 analyzed categories.

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Spanish	Catalan	Euskera
Correction	4.60	2.67	3.27	3.80	0.80	10.00*	8.87	7.53
Fluency	4.47	2.73	3.13	3.67	0.53	10.00*	8.40	6.93
Coherence	4.53	2.60	3.07	3.67	0.33	10.00*	8.47	7.93
Consistency	4.53	2.73	2.87	3.67	0.00	10.00*	8.73	7.93

* Spanish shows a score of 10 because it has been the reference language when making comparisons.

Average language evaluation.



Comprehension evaluation: **Understanding of tasks by AI is very deficient when expressed in indigenous languages (2.3 out of 10)**

The capacity for abstraction is the most affected comprehension trait on the first attempt, scoring 1.9 out of 10.

> Faces serious limitations in understanding sequential context, cultural scope, and humor, especially in visual or technical formats.

Executive evaluation. Breakdown of the 4 analyzed categories.

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Spanish	Catalan	Euskera
Precision	4.47	2.93	2.53	3.73	0.20	10.00*	8.00	6.13
Completeness	4.47	2.93	2.33	3.73	0.40	10.00*	8.47	6.07
Abstraction	4.53	2.87	2.60	3.80	0.87	10.00*	6.73	6.07
Understanding	4.47	2.80	2.47	3.80	0.27	10.00*	8.87	5.87

* Spanish shows a score of 10 because it has been the reference language when making comparisons.

Reflected culture evaluation: Western cultural traits predominate when interacting in indigenous languages (1.7 out of 10)

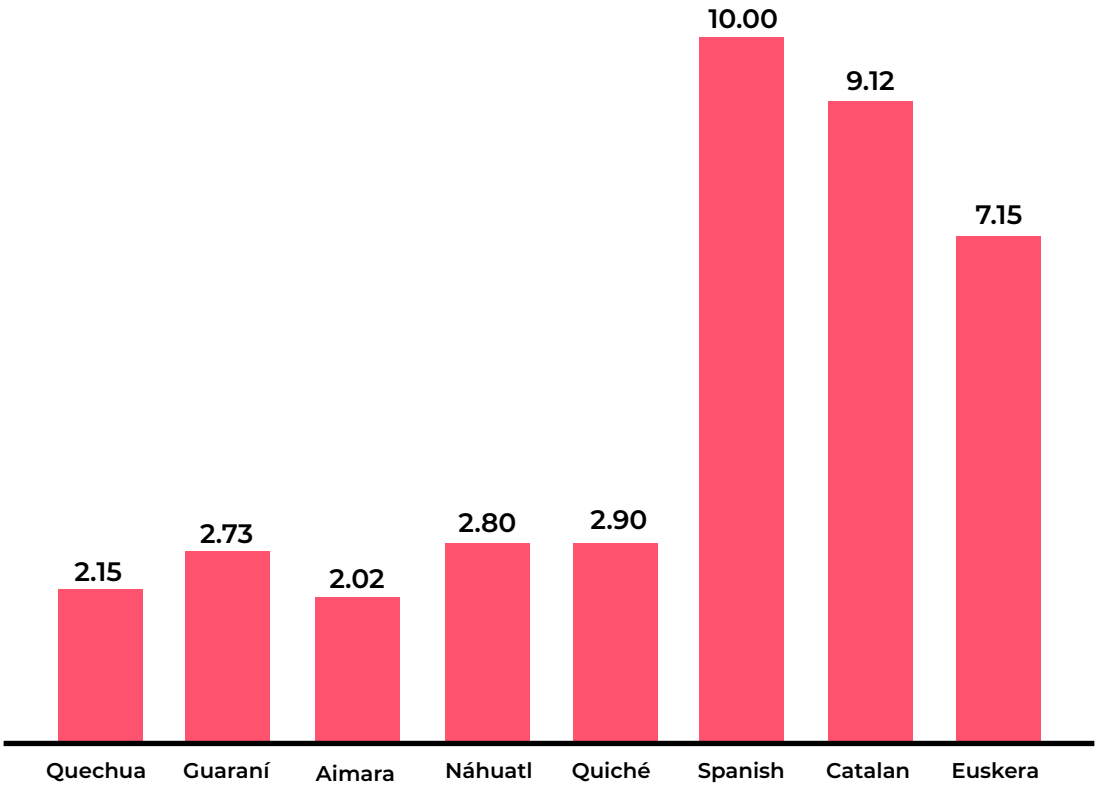
Regarding cultural bias, even the best indigenous language (Quechua) scores below 2.3 out of 10.

Behavioral evaluation. Breakdown of the analyzed categories.

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Spanish	Catalan	Euskera
Cultural bias	2.20	2.27	2.47	0.93	0.80	10.00*	9.20	8.00
Suitability	2.13	3.67	1.80	4.67	5.33	10.00*	9.33	6.40
Adaptation	2.13	2.53	1.40	4.67	4.67	10.00*	9.00	6.20

* Spanish shows a score of 10 because it has been the reference language when making comparisons.

Average behavioral evaluation.



2. IMPACT ON COMMUNITY AND MARKET

AI represents a great opportunity to reduce cultural isolation and give visibility to indigenous peoples and cultures

a **AI is a platform to give visibility to indigenous culture and language**

Conversational AI systems are among the fastest-growing services on the internet.

ChatGPT alone has over 300 million weekly users and receives more than 1 billion questions per day.

Proper positioning of indigenous language and culture in AI increases its potential traditional reach.

b **Automatic translation by AI can help alleviate healthcare access problems for indigenous communities**

In Quechua communities with infant mortality of 44 per 1,000 live births, and Guaraní communities with 16.9 deaths per 1,000 inhabitants.

Telemedicine with AI would allow faster and more effective care, improving health without requiring travel.

c **AI promotes inclusive public administration for indigenous communities**

Multilingual AI systems in government portals and public service applications allow transactions, access to subsidies and social programs without intermediaries.

This significantly increases indigenous participation in state programs.

d **Automating the generation and understanding of indigenous languages will help preserve them**

40% of the world's languages are in danger of extinction, and less than 2% have a presence on the internet.

AI can help preserve them through automatic translators, voice assistants, and educational tools.

This facilitates their use in digital environments and reduces the cultural and technological gap.

AI also represents a significant threat if its training with indigenous knowledge and language is not addressed with determination

a

AI ineffective in indigenous languages increases the gap and exclusion of monolingual and non-literate populations

Although indigenous monolingualism has decreased in recent years, it still represents a relevant segment of indigenous communities, especially women and children.

Inefficient AI in these languages not only creates a gap between cultures but also between genders, perpetuating role differences within indigenous communities.

b

Westernized AI poses a threat to the preservation of indigenous culture

AI models, when trained with predominantly Western data, may lead to misrepresentations and marginalizations*.

This reveals economic discrimination (assumption of poverty, 100%), ruralization (even in areas where more than 35% of the population is urban) and exoticization (clothing, settings,etc.).

c

Westernized AI poses a threat to the preservation of indigenous culture

Many indigenous languages contain unique knowledge about the environment, traditional medicine, and worldview.

Unfortunately, 38.4% of indigenous languages in Latin America (of the 556) are at risk, 18% more compared to 2009.

AI and digital content are the linguistic diffusion channels of the future. Not leveraging them condemns the transmission of these cultures.

d

Poor AI in indigenous languages distanced its communities from benefiting from the economic growth AI will generate

By 2030, AI is estimated to facilitate the creation of 170M new jobs and the loss of 92M

AI is also estimated to represent 3.5% of world GDP.

In the education sector, AI is estimated to grow at a Compound Annual Growth Rate of over 10% between 2023 and 2032.

Online education is booming, with an expected increase of €111,000M between 2024 and 2028.

* Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach (https://arxiv.org/abs/2407.14779?)

3. THE ENABLING ENVIRONMENT: KEY PROGRAMS AND PLAYERS

National fragmentation of support programs limits the scope of support for indigenous languages and cultures

Closer collaboration between national governments and international programs is a strategic opportunity to explore.

1. National support programs for indigenous languages

- The most abundant in number.
- More focused on regional dialect than on the matrix language.
- Accentuate fragmentation.
- Mostly focused on learning culture and language.

2. International programs and NGOs

- Less abundant than national ones.
- Promote cooperation between countries, organizations, and indigenous communities to revitalize and preserve endangered languages.

3. Private companies and indigenous languages

- Tech firms leads the main initiatives supporting indigenous languages in AI.
- B2C companies in the region drive indigenous inclusion programs.
- Both factors represent an important development opportunity for tech SMEs in AI.

The highest impact indigenous inclusion initiatives are those that involve joint participation of government programs, international programs, NGOs, and private companies.

National programs typically focus on just one dialect of the language

Having a minimum consensus dialect would increase interest in languages that bring together a larger number of speakers.

1. National support programs for indigenous languages

- The most abundant in number.
- More focused on regional dialect than on the matrix language.
- Accentuate fragmentation.
- Mostly focused on learning culture and language.

Country	Initiative	Description
Peru	Free Central Quechua Course	Online course to strengthen cultural identity and linguistic rights.
Paraguay	Extended School Day Project (JEE)	Communicative teaching in 300 schools, teacher training, and audiovisual resources.
Chile	Languages are the Future	Immersion workshops and seminars to revitalize and spread the use of indigenous languages.
Bolivia	Educa Bolivia / Wiñay Aru Course	Plurilingual educational programs and virtual course to promote Aimara.
Guatemala	Academy of Mayan Languages / 'Aprendo en Casa y en Clase' Portal	Educational materials, teacher training, and self-learning guides in Mayan languages.
Mexico	Projects of the National Institute of Indigenous Peoples (INPI)	Intercultural bilingual education programs, teacher training, and educational materials.

International programs are often so transversal that they are more difficult to activate in concrete actions

They usually establish strong links between cultural and environmental preservation.

2. International programs and NGOs

- Less abundant than national ones.
- Promote cooperation between countries, organizations, and indigenous communities to revitalize and preserve endangered languages.

Initiative	Description	Reach
Rising Voices	Supports indigenous digital creators in revitalizing and promoting indigenous languages through digital technologies.	Global
IIALI	Promotes the use, conservation, and development of indigenous languages.	Latin America and the Caribbean
FILAC	Promotes cultural preservation and social cohesion through strengthening indigenous languages.	Latin America and the Caribbean
Cultural Survival	Defends the rights of indigenous peoples, supporting self-determination and the preservation of their languages and cultures since 1972.	Global
TACT	Collaborates with indigenous communities in the Amazon to protect tropical forests, biodiversity, and strengthen traditional culture.	Amazon region
IWGIA	Works in defense of indigenous peoples' rights and develops the "Indigenous Navigator" initiative to collect data.	Global

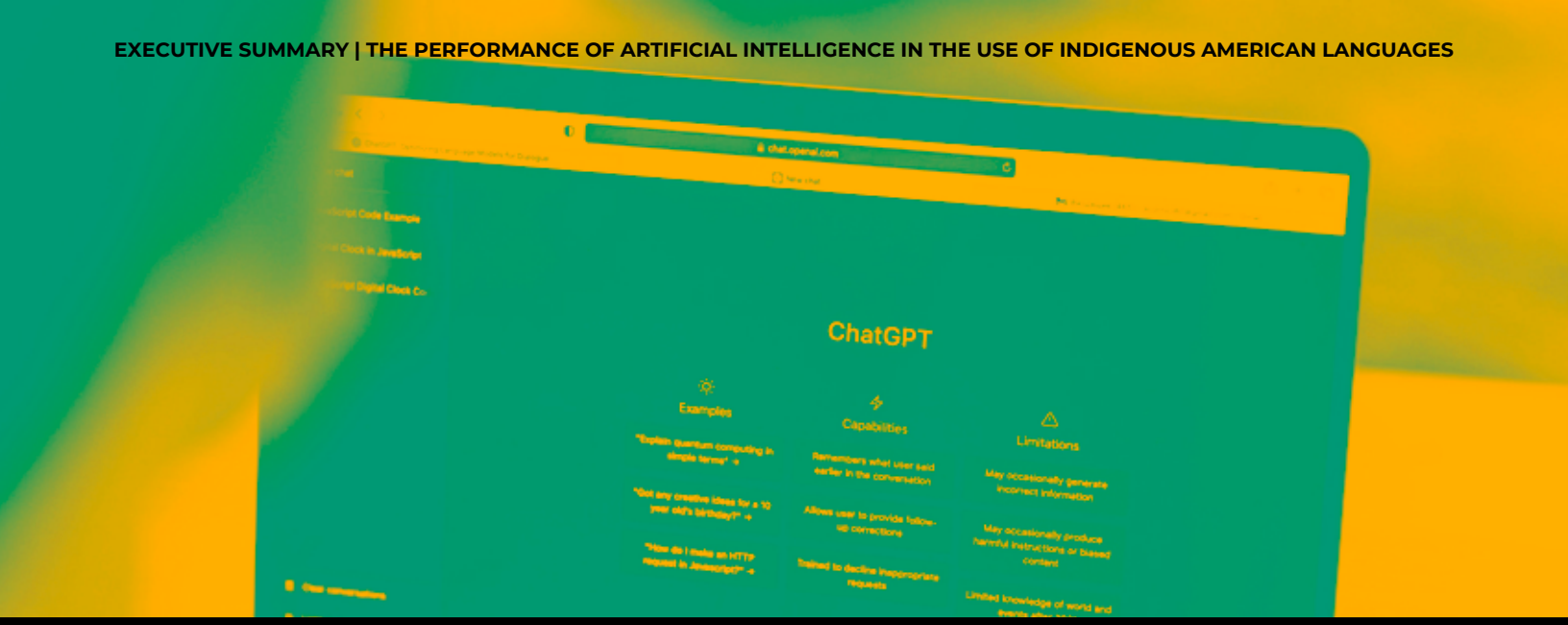
Tech firms companies lead the main initiatives for AI development in indigenous languages

B2C companies can represent an activation opportunity if they make the leap from digitization to AI in indigenous languages

3. Private companies and indigenous languages

- Tech firms leads the main initiatives supporting indigenous languages in AI
- B2C companies in the region drive indigenous inclusion programs.
- Both factors represent an important development opportunity for tech SMEs in AI.

Company	Initiative/Language	Description
Microsoft	AI for Good Lab	Based on AI applications developed around sustainability, humanitarian action, and health.
Google	Woolaroo	Application to describe images in 17 endangered languages.
Microsoft	Quechua chanka	Windows and Office 365 versions in Quechua chanka in collaboration with the Ministry of Culture.
Google	Náhuatl	Incorporation of Náhuatl in Google Translate.
BCP	Quechua	Implementation of Quechua in more than 2,300 ATMs.
BBVA	Quechua	Inclusion of Quechua in educational episodes through an alliance with TV Peru.
CEMEX	Náhuatl	Production of educational capsules in native languages to promote community resilience.
Banrural	Maya Quiché	Installation of multilingual ATMs with operations in native languages.



4. WHAT AN AI NEEDS TO SPEAK A LANGUAGE WELL

There is a very high correlation between the volume of open content in a language and the performance of AI in that language

	Quechua	Guaraní	Aimara	Náhuatl	Quiché	Catalan	Euskera	Spanish
N° of posts on X (x1000)	7,231	985	2,387	1,650	375	92,850	2,912	5,531,599
N° urls in Common Crawl	0.0006%	0.0008%	0.0001%	0	0	0.20%	0.04%	4.62%
N° of Wikipedia editors	40	30	15	14	0	1,503	751	14,171
N° of Wikipedia entries	24,073	5,828	5,186	4,312	0	764,108	448,156	1,992,728
AI performance	3.72	2.77	2.53	3.42	1.25	8.58	6.92	10

Analyzed AIs divided into ownership categories

Proprietary models performance	4.35	3.35	3.05	4.89	1.85	10.00	9.56	8.66
Open models performance	2.78	1.89	1.75	1.22	0.35	10.00	7.11	4.32

Correlation of data volume with respect to AI performance

0.91 0.90 0.80 0.73

0.87 0.86 0.75 0.66

Regarding the performance of proprietary models

0.93 0.94 0.88 0.84

Regarding the performance of open-weights models

The factor that most determines the performance of an AI in a language is the amount of data in that language with which it has been trained

LLMs do not use language-dependent techniques in their training process.

AI performance and available data

84%

correlation between data and performance

- The more data, the better the AI performance.
- This effect is especially pronounced in open source models.

Voice-to-text technologies would allow a very significant improvement

- Much of indigenous knowledge and culture is orally transmitted.
- Having quality voice2text technology increases the possibilities of digitizing content.

Automatic translators are the other great technological enabler

- Having automatic translation technology increases reduces barriers to knowledge and technology access.
- More available knowledge increases the incentive for communities to have a more active digital presence.

The importance of available data in a language for AI to learn to speak it well

The number of entries in Wikipedia is the best predictor of the quality with which AI expresses itself in a language (91% correlation).

Correlation between the volume of data available in a language and the performance of AI in that language.

	N° of posts on X	N° urls in Common Crawl	N° of Wikipedia editors	N° of Wikipedia entries
Performance of all analyzed AIs	0.73	0.80	0.90	0.91
Performance of proprietary models	0.66	0.75	0.86	0.87
Performance of Open Source models	0.84	0.88	0.94	0.93

- **Open source** models (those that offer more possibility of being adapted) show **the greatest dependency on the volume** of data in a language.
- The data suggests that proprietary models may have non-open data sources.

The availability and accuracy of automatic language processing tools also explain AI performance in a language

Automatic language detectors are the automatic processing tool most related to performance.

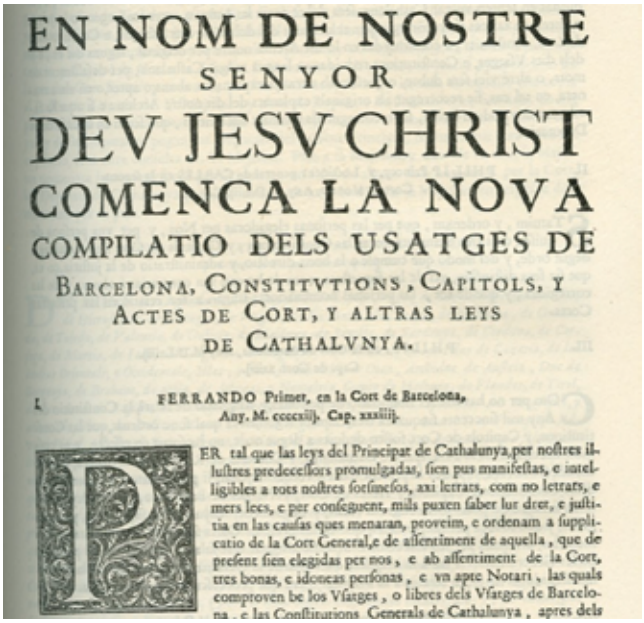
Correlation between automatic processing tools available in a language and the performance of AI in that language.

	Number of automatic translators	Number of automatic language detectors	Text2speech + Speech2text
Language Performance	0.91	0.96	0.92
Executive Performance	0.89	0.92	0.88
Behavioral Performance	0.74	0.89	0.81

The importance of language automatic processing tools, especially speech converters, are especially important in languages with limited written tradition or little digitalization.

A very reduced presence of written tradition in indigenous languages is one of the factors that explains their low performance in AI compared to Western languages with a similar number of speakers

A greater intensity of regional government programs and funds supporting the language, typically linked to nationalist sentiments, is another key that helps explain the phenomenon.

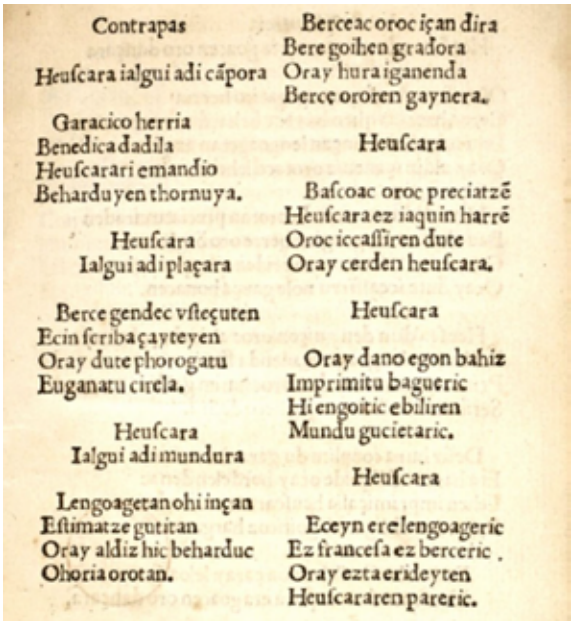


EN NOM DE NOSTRE
SENYOR
DEV JESV CHRIST
COMENCA LA NOVA
COMPILATIO DELS USATGES DE
BARCELONA, CONSTITVTIONS, CAPITOLS, Y
ACTES DE CORT, Y ALTRAS LEYS
DE CATHALVNYA.

FERRANDO Primer, en la Cort de Barcelona,
Any. M. cccxiij. Cap. xxxiij.

PER tal que las leys del Principat de Cathalunya per nostres illustres predecessors promulgadas, sien pus manifestas, e intelligibles a tots nostres soſmesos, axi letrats, com no letrats, e mers lecs, e per conseguent, mils puxen saber lur dret, e justitia en las causas ques menaran, proveim, e ordenam a supplicatio de la Cort General, e de assentiment de aquella, que de present sien elegidas per nos, e ab assentiment de la Cort, tres bonas, e idoneas personas, e vn apte Notari, las quals comproven be los Vsatges, o libres dels Vsatges de Barcelona, e las Constitutions Generals de Cathalunya, apres dels

Compilation of the Usatges of Barcelona, and the Catalan Constitutions (1413) (printed edition).



Contrapas
Heuſcara ialgui adi cãpora
Garacico herria
Benedica dadila
Heuſcarari emandio
Behardu yen thornuya.
Heuſcara
Ialgui adi plaçara
Berce gendec vſteçuten
Ecin ſcribaçayteyen
Oray dute phorogatu
Euganatu cirela.
Heuſcara
Ialgui adi mundura
Lengoagetan ohi inçan
Eſtimatze gutitan
Oray aldiz hic beharduc
Ohoria orotan.

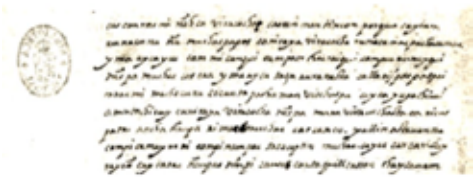
Berceac oroc içan dira
Bere goihen gradora
Oray hura iganenda
Berce ororen gaynera.
Heuſcara
Baſcoac oroc preciatzẽ
Heuſcara ez iaquin harrẽ
Oroc iccaſſiren dute
Oray cerden heuſcara.
Heuſcara
Oray dano egon bahiz
Imprimitu bagueric
Hi engoitic ebiliren
Mundu gucietaric.
Heuſcara
Eceyn ere lengoageric
Ez franceſa ez berceric.
Oray ezta erideyten
Heuſcararen pareric.

Text from the first book printed entirely in Basque by Bernat Etxepare in Bordeaux in 1545.

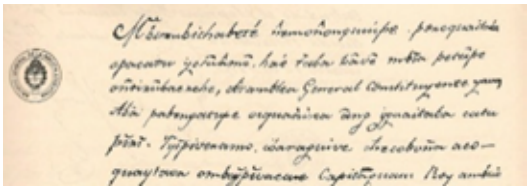
The oldest known texts in Euskera are the Glosas de San Milán from the 11th century For Catalan, fragments of the Forum ludicum and the Les Homilies d'Organyà sermons from the 12th century.

This dissemination was quickly adapted to printing (1440), improving its diffusion.

The first writings in Quechua date from 1584, but it takes 300-500 years to observe writings in certain indigenous languages, which remain isolated and marginal (oral tradition is maintained).



Text from the Huarochirí manuscript in Quechua, 17th.

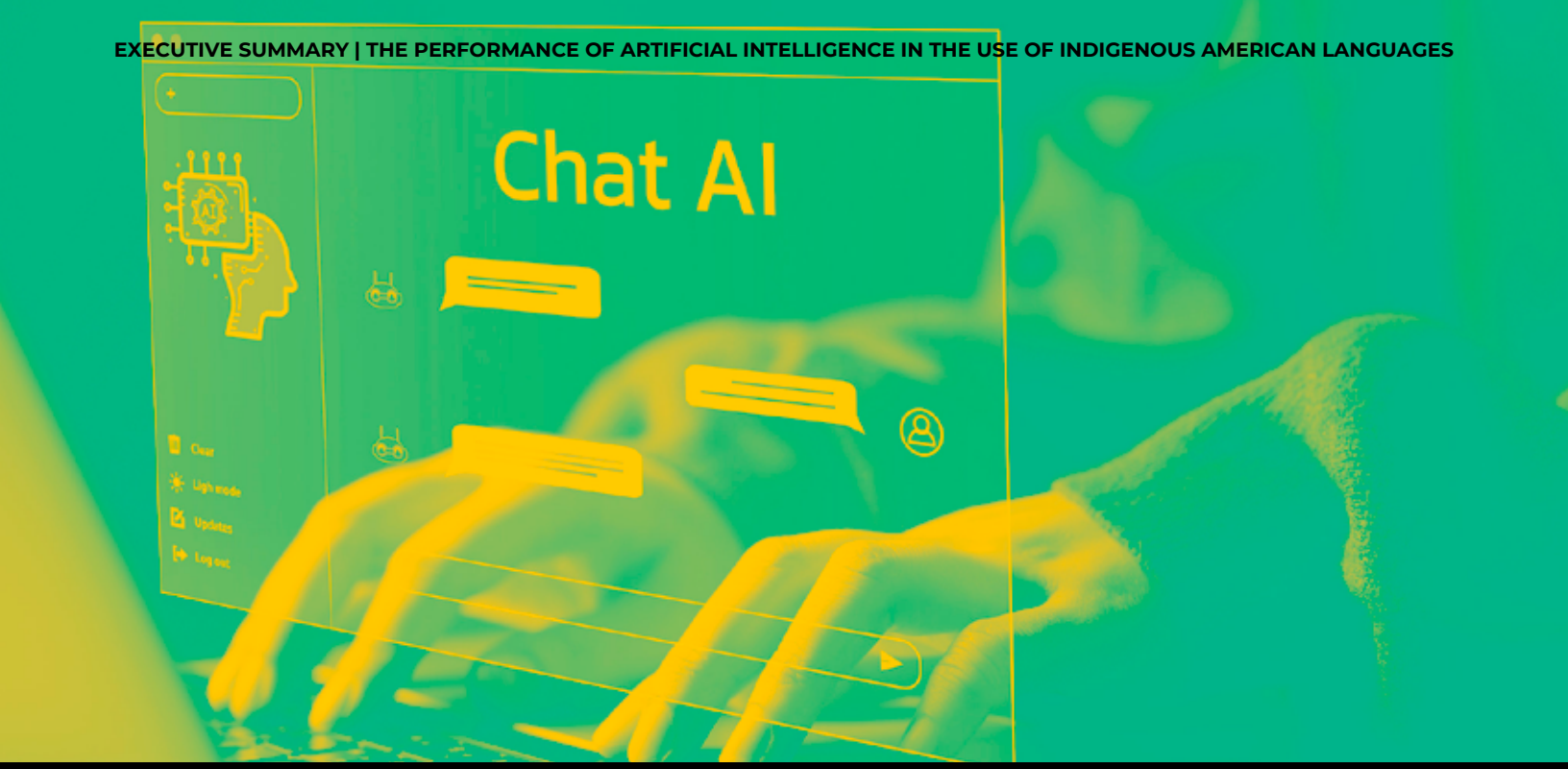


Texts of political modernity in Guarani (1801-1813).

Technology-focused projects that seek to produce improvement results in AI require funds on the order of millions of dollars

And their results have no direct impact on the performance of the most globally used AIs.

Project	Language	Budget (€)	Objective
NEL-AINA	Catalan	**3.000.000**	Generate corpus and computational models of the Catalan language so that companies creating applications based on artificial intelligence (AI), such as voice assistants, Internet search engines, automatic translators and correctors, conversational agents, among others, can easily do so in Catalan.
NEL-GAITU	Euskera	**2.000.000**	Develop and offer basic and cross-cutting linguistic services to use them in all public administrations and offer better public services to citizens.
NÓS	Galician	**2.000.000**	Create digital and linguistic resources necessary to facilitate the development of applications based on artificial intelligence (AI) and language technologies (LT) such as voice assistants, automatic translators and conversational agents in Galician.



5. 21 TECHNOLOGICAL INCLUSION STRATEGIES

1 Promote digital communication in indigenous languages

a Promotion of influencers and creators in indigenous languages

- Content creators in indigenous languages generate global visibility and attract new generations.
- Campaigns and collaborations on digital platforms can expand their reach and impact.

b Training in social media through literacy programs

- Include internet and digital media training in existing literacy initiatives.
- Practical workshops teaching how to create and share digital content, encouraging constant production in their languages.

c Development of thematic forums and platforms

- Creating digital spaces to collect knowledge and traditions preserves cultural heritage and stimulates its use in modern contexts.
- These initiatives promote intergenerational collaboration, drive digital inclusion, and strengthen linguistic pride.

2 Preserve and expand existing content in indigenous languages

a Support programs for digital resource maintenance

- Many digital resources in indigenous languages have been lost due to lack of technical support and expired domains.
- It is essential to create permanent hosting programs and technical support to preserve these contents long-term.

b Digitization and conservation of physical archives

- Digitizing physical archives preserves and expands access to indigenous cultural and linguistic knowledge.
- Using standard formats and public repositories improves integration and visibility in international databases.

c Access to funds for documentation projects

- Collaborations between institutions, NGOs, and governments can secure funds for documentation and digitization of indigenous languages.
- Funding should include technical and cultural aspects, ensuring the continuity and evolution of these contents.

3 Normalize the use of indigenous languages and reduce fragmentation

a Promote agreements for dialectal convergence

- A common standard in indigenous languages improves the effectiveness and reach of resources such as dictionaries, applications, and digital corpora.
- Dialectal standardization facilitates the use of languages in education, media, and formal contexts.

b Respect diversity within a flexible standardization framework

- Standards should be flexible, adapting to different contexts and respecting linguistic diversity.
- Customizable digital resources help balance unification with the preservation of local expressions.

c Promote the training of local experts in linguistics and technology

- Training local experts in linguistics and technology is key to leading and maintaining the standardization of their languages.
- Linguistic normalization reduces fragmentation, improves technological inclusion, and enhances digital content production.

4 Drive the development of enabling tools

a **Development of voice-to-text conversion system**

- Allow increasing the size of trainable data by converting audio or videos to text.
- Facilitate technology access for non-literate indigenous language speakers.

b **Development of automatic detectors for indigenous languages**

- Classify and filter texts by language to manage massive data.
- Improve corpus creation and AI model training.

c **Creation and maintenance of automatic translators**

- Facilitate access to and learning of indigenous languages, reducing linguistic barriers.
- Increase the value proposition of the internet for indigenous communities.

5 Leverage linguistic inclusion initiatives of large companies

a **Alliances with Tech firms and large companies with inclusion initiatives**

- Take advantage of interface localization initiatives to propose a further step with AI incorporation.
- Increase collaboration with technology giants that manufacture universal access technologies.

b **Development of conversational AI technologies**

- Virtual assistants in indigenous languages improve accessibility and user experience.
- Drive the need for improvement in AI performance in indigenous languages.

c **Generate demand for resources and training**

- The incorporation of indigenous languages in popular services creates positive pressure on the technological ecosystem.
- The more services are used in native languages, the greater the incentive to continue developing advanced technologies.

6 Expand connectivity in indigenous communities

a Promote internet coverage expansion programs

- Increasing internet access in indigenous communities implies increasing the generating collective in indigenous languages.
- The increase in the network effect increases the incentive for content generation action.

b Accompany with training programs in internet use

- Implement digital training adapted to local languages and cultures.
- Teach navigation, security, content creation, and access to key resources.

c Extend the reach of existing literacy programs

- Integrate digital training into literacy programs in indigenous languages.
- Encourage participation in digital platforms to generate content in native languages.

7 Increase linguistic localization of tech services

a Localization of operating systems and web browsers

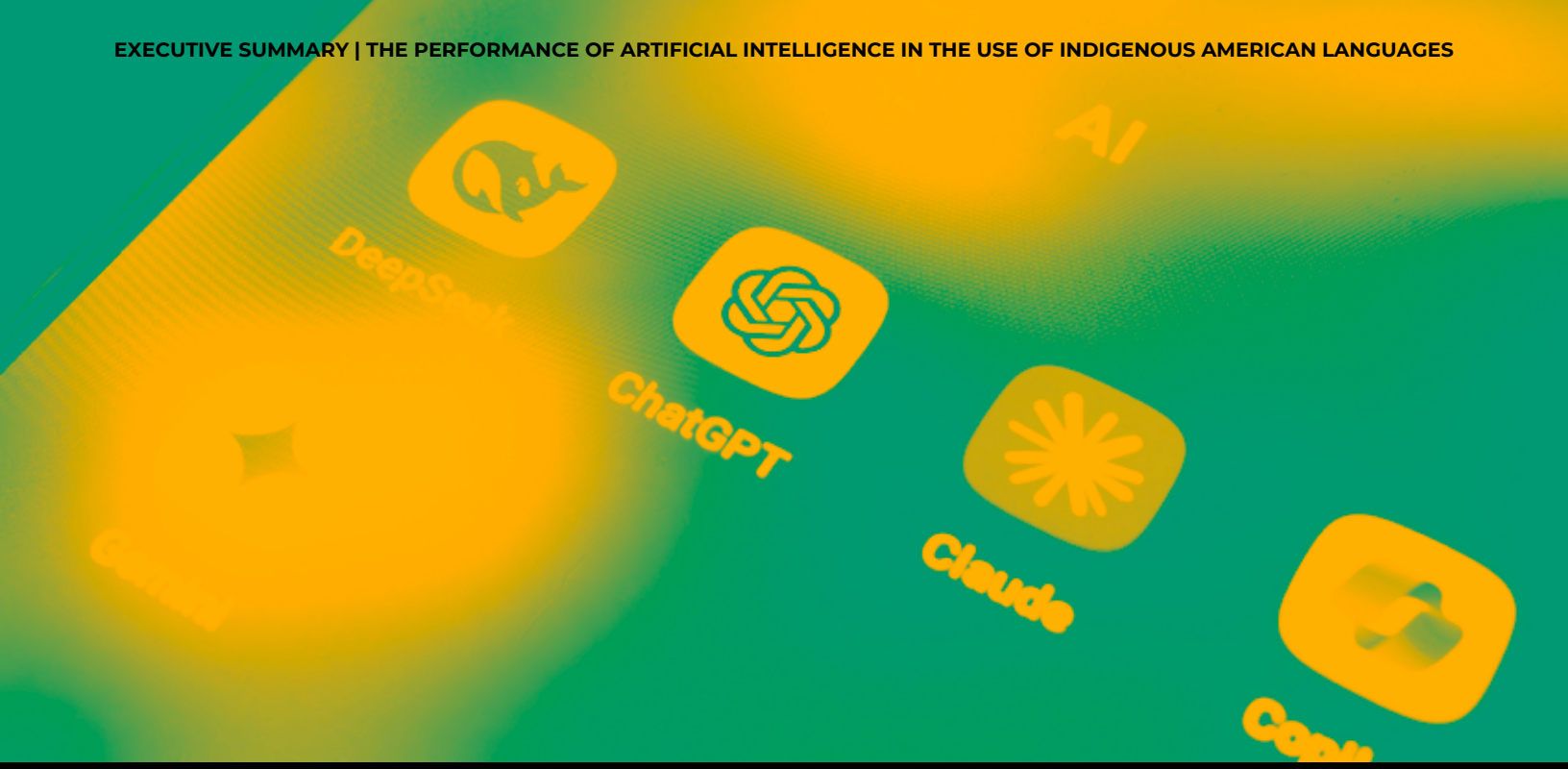
- Localizing operating systems and browsers in indigenous languages increases community access.
- Collaborate with technology companies for equitable translation and customization.

b Localization of search engines and social networks

- The translation of interfaces in social networks increases the possibility of generating conversations in indigenous languages.
- Providing tools in these languages facilitates access to relevant cultural content.

c Awareness and collaboration with big tech companies

- Demonstrations of social impact and studies on the digital divide to persuade them to expand their commitment.
- Increasing localization promotes digital equity and cultural recognition.



6. RECOMMENDATIONS AND ACTION PLAN

1 Creation of an international consortium to drive the project

The first step is to form a consortium composed of national and international organizations, institutions dedicated to cultural protection, and technology companies interested in linguistic inclusion. This consortium will aim to lead and coordinate all strategic actions of this plan.

a Consortium responsibilities

- Definition of general strategy and supervision of the action plan.
- Management of alliances and financing.
- Coordination between the different actors involved.
- Evaluation of results and adjustments to the plan.

2 Creation of the Implementation Working Team

This team will be responsible for executing the defined strategic actions, organizing activities, and ensuring the correct implementation of the plan. It will be composed of members from the organizations that have constituted the project driving consortium.

a IWT team functions

- Establish and develop strategic alliances with sponsors of local initiatives.
- Supervise the execution of projects derived from strategies.
- Manage assigned resources.
- Guarantee the active participation of indigenous communities.

3 Organization of a high-visibility event to communicate the initiative

An event will be organized to present the study results, communicate the formation of the consortium, and announce the strategic alliances achieved. It will also be used as a platform to convene a high-impact hackathon.

a Event objectives

- Present the report on the AI gap in indigenous languages.
- Announce the creation of the consortium.
- Present the roadmap.
- Invite the community to participate in the hackathon.

4 Technological Innovation Hackathon for AI in indigenous languages

The hackathon will bring together key actors, community groups, developers, technology companies, and content creators to design and present concrete projects that contribute to technological inclusion strategies.

a Hackathon objectives

- Break down inclusion strategies into specific initiatives.
- Develop project proposals in collaboration with local communities.
- Promote the creation of technological tools in indigenous languages.

b Hackathon structure

- **Working Groups:** Teams will develop project plans focused on inclusion strategies.
- **Project Presentation:** Projects will be presented to a jury composed of consortium members and sponsors.
- **Evaluation and Awards:** Prizes will be awarded to the most outstanding projects.
- **Roadmap:** Selected projects will form part of the main working team's roadmap.

5 Development of local strategic alliances

It is necessary to identify and establish collaborations with public and private actors, NGOs, technology companies, universities, and media that can act as sponsors of specific initiatives.

a Priority alliances

- Private consumer companies with ongoing initiatives for digitalizing the experience in indigenous languages.
- Local governments that are promoting support programs for indigenous languages.
- Technology providers of large consumer private companies.
- Activist organizations and volunteers supporting indigenous communities.

6 Execution of local projects and monitoring of initiative progress

The execution of local projects will be a priority to achieve direct impact in communities. Teams responsible for each project must coordinate with the ETI to ensure compliance with objectives, deadlines, and the correct allocation of resources. Continuous supervision mechanisms will be established to guarantee the quality of results and the sustainability of initiatives.

a Execution phases

- **Resource Allocation:** Distribution of funds, technology, and training to local teams.
- **Implementation:** Practical development of strategies agreed upon in the hackathon and work plans.
- **Continuous Supervision:** The main team will monitor progress, collecting information on results, obstacles, and good practices.
- **Periodic Reports:** Project managers will present progress reports to the consortium.
- **Evaluation of Results:** Impact will be analyzed in terms of improved AI performance, generation of content in indigenous languages, and community participation.
- **Adjustments and Improvement:** Evaluated results will allow making adjustments to projects in execution, optimizing applied strategies.

IDB
LAB