Letter from the Executive Board

Distinguished Delegates,

It is with great honor that we welcome you to this session of the Global Partnership on Artificial Intelligence (GPAI). We convene at a critical juncture, where the rapid advancement of artificial intelligence presents both unprecedented opportunities and profound challenges to the global community. The topic before this committee—"Deliberating on the role of AI in revitalizing endangered languages and oral traditions, and the need to center native communities in such digital effort"—lies at the very heart of GPAI's mission. Our partnership was founded to guide the responsible development and use of AI, grounded in the principles of human rights, inclusion, diversity, and societal benefit. This debate will test our collective commitment to these values.

The world is currently experiencing a linguistic crisis of staggering proportions. According to UNESCO, a language disappears every two weeks, taking with it an irreplaceable repository of culture, history, and unique human understanding. This is not a passive or natural phenomenon; it is the result of historical and ongoing pressures, including colonization, economic marginalization, and political repression, that have silenced voices and erased heritage. The loss of these languages and the oral traditions they carry represents a catastrophic depletion of what the United Nations has termed the "common heritage of humankind".

Into this crisis enters artificial intelligence, a technology with the potential to document, teach, and create new digital domains for these threatened languages on a scale previously unimaginable. AI-powered tools can transcribe the voices of the last fluent elders, create interactive learning platforms for new generations, and give digital life to languages that have been excluded from the modern world. The promise is that of a digital renaissance, a technological lifeline for cultures on the brink.

However, this promise is shadowed by a significant peril. The very models that power AI are built upon a paradigm of vast data extraction, often without the knowledge, consent, or benefit of the communities from which the data originates. Unchecked, AI risks becoming a new and insidious tool of "digital colonialism," perpetuating historical patterns of exploitation, misrepresenting sacred cultural knowledge, and further marginalizing the very communities it purports to help.

This committee is therefore charged with a task of immense importance. Your deliberations must move beyond a simple technical assessment of AI's capabilities. You are called upon to forge a global policy framework that addresses this fundamental tension. The central question is not *if* AI can be used, but *how* it must be used. The answer, we propose, lies in unequivocally centering the rights, leadership, and sovereignty of Indigenous and native communities.

Your work will require a nuanced understanding of international human rights law, particularly the United Nations Declaration on the Rights of Indigenous Peoples (UNDRIP), and the principles of Indigenous Data Sovereignty. You must explore and advocate for models of co-design, ethical partnership, and community-

led governance. The challenge is to ensure that AI serves as a tool for empowerment and self-determination, not as a vector for appropriation.

We expect a debate that is both ambitious and pragmatic, grounded in evidence and guided by a firm commitment to justice. The resolutions you produce have the potential to shape the future of AI development, ensuring that this powerful technology is harnessed not only to preserve the diversity of human expression but to honor the rights and dignity of its keepers. We look forward to your thoughtful contributions and a productive session of debate.

Sincerely,

The Dais

The Nature of Proof/Evidence in Council

Evidence or proof is acceptable from the following sources

→ News Sources:

State operated News Agencies – These reports can be used in the support of or against the State that owns the News Agency. These reports, if credible or substantial enough, can be used in support of or against any Country as such but in that situation, they can be denied by any other country in the council. Some examples are –

1. IRNA (Iran) http://www.irna.ir/ENIndex.htm,
2. BBC (United Kingdom) http://www.bbc.co.uk/
3. Xinhua News Agency and CCTV (P.R. Of China) http://cctvnews.cntv.cn/

→ Government Reports:

These reports can be used in a similar way as the State Operated News Agencies reports and can, in all circumstances, be denied by another country. However, a nuance is that a report that is being denied by a certain country can still be accepted by the Executive Board as credible information.

Examples are Government Websites like:

1. State Department of the United States of America: http://www.state.gov/index.htm ,
2. Ministry of Defense of the Russian Federation: http://www.eng.mil.ru/en/index.htm ,
3. Permanent Representatives to the United Nations Reports: http://www.un.org/en/members/ (Click on any country to get the website of the Office of its Permanent Representative.)
4. Multilateral Organizations like the NATO (http://www.nato.int/cps/en/natolive/index.htm) ASEAN (http://www.aseansec.org/), OPEC (http://www.opec.org/opec_web/en/), etc.

→ UN Reports:

All UN Reports are considered as credible information or evidence for the Executive Board of the General Assembly.

1. UN Bodies: Like the SC (http://www.un.org/Docs/sc/), GA (http://www.un.org/en/ga/), HRC (http://www.ohchr.org/EN/HRBodies/HRC/Pages/HRCIndex.aspx) etc.
2. UN Affiliated bodies like the International Atomic Energy Agency (http://www.iaea.org/), World Bank (http://www.worldbank.org/), International Monetary Fund (http://www.imf.org/external/index.htm), International Committee of the Red Cross (http://www.icrc.org/eng/index.jsp, etc.
3. Treaty Based Bodies like the Antarctic Treaty System (http://www.ats.aq/e/ats.htm), the International Criminal Court (http://www.icc-cpi.int/Menus/ICC) .

Introduction to Committee

The Global Partnership on Artificial Intelligence (GPAI) is a multi-stakeholder initiative launched in 2020 to guide the responsible development and use of AI in a manner that is grounded in human rights, inclusion, diversity, innovation, and economic growth. Its mission is to bridge the gap between theory and practice on AI by supporting cutting-edge research and applied activities on AI-related priorities.

GPAI brings together leading experts from science, industry, civil society, international organizations, and government who collaborate across four key themes: Responsible AI, Data Governance, The Future of Work, and Innovation & Commercialization. As a forum that fosters international cooperation, GPAI provides a platform for member states to share best practices, identify challenges, and develop a common understanding of the complex issues surrounding AI. Unlike other international bodies that may focus on binding treaties, GPAI emphasizes practical projects, pilot programs, and expert-led analysis to inform policy-making at the national and international levels. Delegates in this committee should focus on creating practical, ethical, and implementable frameworks that align with GPAI's core mission of promoting responsible AI for the benefit of all humanity.

I. Introduction

Language is more than a means of communication; it is the lifeblood of a culture. It encodes a community's unique understanding of the world, its history, values, and ancestral knowledge. Oral traditions are the primary vehicle through which this cultural richness is transmitted across generations. However, globalization, political pressures, and socio-economic marginalization have accelerated the decline of linguistic diversity. The United Nations estimates that a language dies every two weeks, and over 40% of the world's languages are currently endangered, many of them spoken by Indigenous peoples.

The disappearance of a language is an irreversible loss for humanity. It erases unique knowledge systems, including traditional ecological knowledge, medicinal practices, and philosophical outlooks. Revitalizing these languages is therefore a critical task for cultural preservation and human rights.

In this context, Artificial Intelligence (AI) has emerged as a potential game-changer. AI technologies, such as Natural Language Processing (NLP), speech recognition, and machine translation, offer powerful tools for language documentation, education, and revitalization. For instance:

- **AI-powered apps** can create interactive and accessible language-learning platforms for endangered languages that often lack educational resources.
- **Speech-to-text and text-to-speech systems** can help document and transcribe vast archives of oral histories from the last remaining native speakers.
- **Machine translation** can help bridge the communication gap between generations and make content available in native languages, increasing their relevance in the digital age.

However, the application of AI in this sensitive domain is not without significant risks. Most mainstream AI models are trained on massive datasets from dominant languages like English and Mandarin, leaving "low-resource" languages far behind. This creates a risk of **algorithmic bias**, where AI tools may fail to accurately capture the nuances of a language or may even perpetuate harmful stereotypes.

Furthermore, the process of collecting language data raises critical questions about **data sovereignty and intellectual property**. Who owns the language data collected from a community? How can communities control how their cultural heritage is used and commercialized by tech companies? There is a profound danger that well-intentioned efforts could turn into a new form of "digital colonialism," where external entities extract and profit from indigenous knowledge without equitable benefit or consent.

Therefore, the core of this debate is not just about *what* technology can do, but *how* it should be done. The principle of **Free, Prior, and Informed Consent (FPIC)** and the need to center native communities in every stage of the process—from project design and data collection to tool development and implementation—is paramount. The challenge for this committee is to develop a global framework that unlocks the potential of

AI for language revitalization while establishing robust ethical safeguards that empower, rather than exploit, the very communities these efforts are meant to serve.

II. A World of Silenced Voices: The Crisis of Language Endangerment

A. The Scale of the Crisis

The diversity of human language, a cornerstone of our collective cultural heritage, is disappearing at an alarming rate. Of the estimated 7,000 languages spoken in the world today, at least 40% are considered endangered. This figure, while stark, may be a conservative estimate; some projections warn that up to 95% of the world's languages could become extinct or seriously endangered by the end of this century. The pace of this loss is staggering: on average, one language falls silent every two weeks, carrying with it an entire cultural and intellectual universe.

This crisis is exacerbated by a profound digital divide. While globalization has connected the world, it has done so primarily through a handful of dominant languages. Only a few hundred languages have a meaningful presence in educational systems and the public domain, and fewer than one hundred are used in the digital world. This digital exclusion creates a feedback loop: languages without a digital footprint become less relevant in a technologically integrated world, accelerating their decline among younger, digitally native generations. The UNESCO

Atlas of the World's Languages in Danger provides a comprehensive, though sobering, account of this reality, listing approximately 2,500 endangered languages. This includes 577 languages classified as "critically endangered," with only a few elderly speakers remaining, and documents 230 languages that have become extinct since 1950 alone. This data underscores a clear and urgent reality: without concerted international action, a vast portion of human linguistic heritage is on a path to irreversible loss.

B. Defining Endangerment

To effectively address this crisis, it is crucial for delegates to understand that language endangerment is not merely a function of speaker population size. A language's vitality is more accurately measured by its intergenerational transmission—whether it is being passed from elders to children as a living mother tongue. There are languages in Indonesia with as many as two million living native speakers that are considered endangered because all are of advancing age, with little to no transmission to the young. Conversely, the Ladin language in Italy has only 30,000 speakers, but because almost all children still learn it as their first language, it is not currently considered endangered.

This principle of intergenerational transmission is the cornerstone of the framework developed by UNESCO to classify the degrees of endangerment. This framework provides a standardized vocabulary for assessing the health of a language and is essential for prioritizing revitalization efforts.

Table 1: UNESCO's Six Degrees of Language Endangerment

Degree of Endangerment	Intergenerational Language Transmission Description
Safe	Language is spoken by all generations; intergenerational transmission is uninterrupted.
Vulnerable	Most children speak the language, but it may be restricted to certain domains (e.g., home).
Definitely Endangered	Children no longer learn the language as a mother tongue in the home.
Severely Endangered	The language is spoken by grandparents and older generations. While the parent generation may understand it, they do not speak it to children or among themselves.
Critically Endangered	The youngest speakers are grandparents and older, and they speak the language partially and infrequently.
Extinct	There are no speakers left. The Atlas presumes extinction if there have been no known speakers since the 1950s.

Source: Data compiled from UNESCO, 2003.

C. The Drivers of Language Loss

The decline of the world's languages is not a natural, inevitable process. It is the result of a complex interplay of historical, political, and socio-economic forces that actively discourage or prevent communities from using and transmitting their native tongues.

Historical and Political Factors: The legacies of colonialism, imperialism, and nation-state building have been primary drivers of language loss. Policies of forced assimilation have historically targeted Indigenous and minority languages as part of a broader effort to erase distinct cultural identities. This includes genocidal violence, such as that experienced by the native population of Tasmania, and political repression, where states limit or outright prohibit the use of minority languages in schools, media, and public life. The establishment of residential schools in countries like Canada and the United States, where Indigenous children were forcibly removed from their families and punished for speaking their languages, is a stark example of such "linguicide" (language killing).

Socio-Economic Pressures: In the contemporary era, globalization and economic integration have accelerated language shift. The dominance of a few major languages in global commerce, media, and technology creates powerful incentives for individuals to adopt a more "prestigious" or economically advantageous language. This process is often subtle and unconscious; as individuals make daily communication choices to improve their social and economic prospects, they may unintentionally favor a dominant language at the expense of their ancestral one. Urbanization is another key factor, as migration to

cities often necessitates learning the lingua franca of the new environment, leading to a breakdown in language transmission to subsequent generations.

Environmental and Geographic Factors: Research has also identified correlations between environmental conditions and linguistic diversity. Languages with smaller geographic ranges and smaller speaker populations are inherently more vulnerable to demographic and environmental shocks. These conditions are often found in tropical regions with high rainfall and topographic heterogeneity. In contrast, recent rapid language declines observed at higher latitudes have been strongly linked to periods of high economic growth, highlighting the powerful influence of modernization and economic change on linguistic ecosystems.

The common thread among these drivers is power. The endangerment of a language is a symptom of the marginalization of its speakers. Therefore, any viable solution for language revitalization cannot be merely linguistic or technological; it must also address the underlying issues of social, political, and economic inequality.

D. The Irreplaceable Value of Oral Traditions

When a language dies, far more than a set of words and grammar is lost. Language is the primary vehicle for what UNESCO defines as intangible cultural heritage—the living practices, knowledge, and expressions that are passed from one generation to the next. Oral traditions, in particular, represent a vast and sophisticated repository of human knowledge.

These traditions are not simply folktales for amusement; they are highly structured systems for educating the young, transmitting history, reinforcing social norms, and articulating a community's entire worldview. They encode complex knowledge about ecology, traditional medicine, sustainable agriculture, celestial navigation, law, and philosophy. For many societies, oral traditions are as reliable and authoritative as written records, serving as the collective memory and identity of a people. The loss of a language severs the link to this knowledge, effectively erasing a unique library of human experience.

Furthermore, a growing body of evidence shows a direct link between linguistic and biological diversity. Regions with high linguistic diversity often correspond to regions of high biodiversity. This is because Indigenous languages encode detailed, place-based ecological knowledge accumulated over millennia. The disappearance of these languages thus represents not only a cultural loss but also an environmental one, as critical knowledge for sustainable management of ecosystems vanishes with them. Preserving linguistic diversity is therefore an ethical imperative, an act of justice that recognizes the inherent value of diverse ways of knowing, remembering, and living meaningfully in the world.

III. Guiding Principles in a Digital Age: International Frameworks and Actions

The intersection of artificial intelligence and endangered languages does not exist in a policy vacuum. A robust set of international principles and frameworks, developed over decades, provides a clear normative

foundation for this committee's deliberations. These instruments establish the fundamental rights of Indigenous peoples to their culture and language and outline the responsibilities of the international community to safeguard our shared human heritage. The task before this committee is not to invent new rights, but to ensure that existing ones are respected and applied in the rapidly evolving digital landscape.

A. The UN Declaration on the Rights of Indigenous Peoples (UNDRIP)

Adopted by the UN General Assembly in 2007, UNDRIP is the most comprehensive international instrument on the rights of Indigenous peoples and serves as the foundational human rights framework for this topic. It affirms that Indigenous peoples are equal to all other peoples and have the right to be free from discrimination. Several articles are of direct relevance to the development and deployment of AI for language revitalization:

- **Article 3: The Right to Self-Determination:** This is the cornerstone of the Declaration. It establishes that "Indigenous peoples have the right to self-determination. By virtue of that right they freely determine their political status and freely pursue their economic, social and cultural development". This principle must extend to the digital realm, meaning Indigenous communities have the right to determine for themselves how, or if, technologies like AI are used in relation to their cultural heritage.
- **Article 13: The Right to Language:** This article explicitly grants Indigenous peoples the right "to revitalize, use, develop and transmit to future generations their histories, languages, oral traditions, philosophies, writing systems and literatures". It further obligates states to take effective measures to ensure this right is protected.
- **Article 31: The Right to Cultural Heritage and Intellectual Property:** This is a critical provision that directly addresses the data-extractive nature of many AI models. It states that Indigenous peoples have the right "to maintain, control, protect and develop their cultural heritage, traditional knowledge and traditional cultural expressions... They also have the right to maintain, control, protect and develop their intellectual property over such cultural heritage, traditional knowledge, and traditional cultural expressions". This assertion of control challenges the practice of scraping cultural and linguistic data without consent to train commercial algorithms.
- **Free, Prior and Informed Consent (FPIC):** Embedded throughout the Declaration, FPIC is a key principle requiring that Indigenous peoples be meaningfully consulted before any action is taken that affects their rights, lands, or resources. In the context of AI, this means that no Indigenous language data should be collected, used, or incorporated into an AI system without the free, prior, and informed consent of the community concerned.

B. UNESCO's Mandate to Safeguard Heritage

The United Nations Educational, Scientific and Cultural Organization (UNESCO) leads international efforts to protect cultural heritage. The 2003 Convention for the Safeguarding of the Intangible Cultural Heritage is a key global framework that recognizes the importance of living traditions.

While the Convention's scope does not include the direct preservation of languages themselves, it is fundamentally relevant to this committee's work. Article 2 of the Convention explicitly identifies "language as a vehicle of the intangible cultural heritage". This recognition is crucial because it acknowledges the inseparable link between a language and the oral traditions it conveys. The death of a language inevitably leads to the permanent loss of the stories, songs, rituals, and knowledge it carries.

Crucially, UNESCO's approach emphasizes that safeguarding is not about ossifying traditions as static "products" in a museum. Instead, it is about supporting the living "processes" through which communities constantly recreate and transmit their heritage. This principle aligns perfectly with the dynamic potential of AI. AI tools should not be seen as creating a final, definitive archive of a language, but rather as supporting the living, evolving use of that language by its community of speakers.

C. The Global Partnership on Artificial Intelligence (GPAI)

This committee, the Global Partnership on Artificial Intelligence, is uniquely positioned to address this topic. Launched in June 2020, GPAI is a multistakeholder initiative that brings together leading experts from science, industry, civil society, and government to bridge the gap between theory and practice in AI. Its founding mission is to "support and guide the responsible adoption of AI that is grounded in human rights, inclusion, diversity, innovation, economic growth and societal benefit". This mandate is built upon the shared values expressed in the OECD Recommendation on Artificial Intelligence.

GPAI's structure, which includes four working groups—Responsible AI, Data Governance, Future of Work, and Innovation & Commercialization—provides a robust framework for tackling this multifaceted issue. The **Data Governance Working Group** is particularly pertinent. Its mandate is to provide expertise "to promote data for AI being collected, used, shared, archived and deleted in ways that are consistent with human rights, inclusion, diversity, innovation, economic growth, and societal benefit".

Previous and ongoing GPAI projects provide a direct foundation for this committee's work. The "Advancing Data Justice" project, for instance, produced guides for policymakers and impacted communities on how to achieve more equitable data governance. More recently, a GPAI Innovation Workshop held in Tokyo in May 2025 featured a dedicated "Multilingual and Multicultural AI" group. This group acknowledged that mainstream AI systems fail to reflect the world's cultural diversity and proposed the establishment of a "Multicultural AI Consortium." This proposed consortium would aim to develop datasets for undervalued languages, utilizing the UN's "Endangered Languages" list, and create benchmarks for evaluating the cultural safety of AI systems.

The convergence of these international frameworks is clear. UNDRIP provides the human rights foundation, UNESCO provides the cultural preservation mandate, and GPAI provides the specific institutional venue for translating these principles into actionable policy for the age of AI. The challenge for this committee is to build upon this existing work, particularly GPAI's own nascent efforts in multicultural AI, to create a comprehensive and effective global strategy.

IV. The Digital Renaissance: AI's Potential and Peril in Language Revitalization

Artificial intelligence offers a new frontier for language revitalization, providing a suite of powerful tools that can accelerate documentation, create novel educational resources, and expand the digital presence of endangered languages. However, these technologies are not neutral. Their design, data requirements, and deployment carry inherent risks that could replicate historical patterns of extraction and misrepresentation. A balanced understanding of both the immense potential and the significant perils is essential for developing responsible policy.

A. A New Generation of Tools: AI Applications

Advances in AI, particularly in Natural Language Processing (NLP), have unlocked a range of applications with direct relevance to language revitalization efforts:

- **Documentation and Archiving:** One of the most urgent tasks in language preservation is documenting the speech of the last remaining fluent elders. This has traditionally resulted in thousands of hours of audio recordings that are difficult to search and use. **Automatic Speech Recognition (ASR)** technology can automate the transcription of these archives, transforming them into searchable, usable digital texts. This not only preserves the knowledge but also creates foundational data for other AI tools.
- **Language Learning and Education:** AI can revolutionize how endangered languages are taught and learned. NLP models can be used to develop a wide array of educational materials, including digital dictionaries, grammar analysis tools, and automated quiz generators.
- **Conversational AI**, in the form of chatbots, can provide learners with invaluable opportunities for low-pressure conversational practice, a critical component of language acquisition that is often limited by the small number of fluent speakers.
- **Content Creation and Accessibility: Text-to-Speech (TTS)** systems can convert written text into spoken language, creating audiobooks, voiced dictionaries, and accessibility tools for learners.
- **Machine Translation (MT)** offers the potential to translate web content, software, and other digital resources into endangered languages, thereby creating a digital ecosystem where the language can be used. However, the accuracy of MT for low-resource languages remains a significant technical hurdle.
- **Overcoming Data Scarcity:** A major barrier for applying AI to most of the world's languages is the lack of large digital datasets required for training. This is known as the "low-resource" problem. Researchers are actively developing techniques to address this.
- **Small Language Models (SLMs)** are a promising approach; these models are smaller and can be trained effectively on more limited, language-specific datasets. Techniques like

- **distillation**, where a smaller model is trained using data generated by a larger one, and **transfer learning**, where a model pre-trained on a high-resource language is fine-tuned for a low-resource one, are helping to make AI tools more feasible for a wider range of languages.

B. Case Studies in AI-Powered Language Revitalization

The application of these technologies is not merely theoretical. Around the world, communities and organizations are pioneering innovative projects that demonstrate both the potential of AI and the importance of community-led governance.

Table 2: Case Studies in AI-Powered Language Revitalization

Project Name	Indigenous Community/Language	Key AI Technology	Governance Model	Key Principles
Te Hiku Media	Māori (Te Reo Māori), New Zealand	Automatic Speech Recognition (ASR)	Indigenous-led non-profit	Indigenous Data Sovereignty, Kaitiakitanga (Guardianship) License
First Languages AI Reality (FLAIR)	North American Indigenous Languages (e.g., Wakashan, Algonquian)	ASR, Voice AI	Indigenous-led initiative	Data Sovereignty, Linguistic Self-Determination, Open Source
Masheli Chatbot	Choctaw Nation (Choctaw), USA	Conversational AI (Chatbot)	Indigenous-led academic research	Community-based development, Low-pressure learning
Google's 1,000 Languages Initiative	Global (various low-resource languages)	Universal Speech Model (USM), MT	Corporate-led initiative	Universal Accessibility, Large-scale modeling
Microsoft & Nunavut Partnership	Inuit (Inuktitut), Canada	Text-to-Speech (TTS)	Corporate-Government Partnership	Digital Accessibility, Language Preservation

The **Te Hiku Media** project in New Zealand is a leading example of an Indigenous-led approach. Rather than ceding control to external platforms, they built their own digital archive, *Whare Kōrero* ("house of speech"), to house over 30 years of recordings. They developed a high-accuracy ASR model for te reo Māori using this data and, crucially, created a bespoke *kaitiakitanga* (guardianship) data license. This license ensures that the community retains ownership and that the data is used only for the benefit of the Māori people, explicitly prohibiting its use for harmful purposes like surveillance.

Similarly, the **First Languages AI Reality (FLAIR)** initiative in North America is founded and led by Indigenous technologists. Their work focuses on developing novel ASR methods that drastically reduce the amount of data required, making the technology accessible to communities with very few speakers or recordings. Their model is explicitly designed to respect data sovereignty and linguistic self-determination, with a commitment to sharing all tools as open-source resources.

These Indigenous-led projects stand in contrast to large-scale corporate initiatives like **Google's "1,000 Languages Initiative"** and **Meta's "No Language Left Behind"**. While these projects bring enormous computational resources and have the potential to rapidly expand language support, they raise fundamental questions about data governance. It is often unclear where the training data is sourced from, whether consent was obtained, and how the communities whose languages are being digitized will benefit from the commercialization of these technologies.

C. Risks of Digital Colonialism and Algorithmic Harm

The enthusiasm for AI's potential must be tempered by a clear-eyed assessment of its significant risks. Without careful, community-centered governance, these technologies can cause profound harm.

- **Data Extraction and Intellectual Property:** The dominant paradigm in AI development involves training models on vast datasets, much of which is scraped from the internet. When this data includes Indigenous languages, traditional stories, or sacred knowledge, it constitutes a form of digital extraction that replicates colonial patterns of resource appropriation. Existing intellectual property laws, with their focus on individual authorship and limited terms of protection, are fundamentally ill-equipped to protect the collective, intergenerational nature of Indigenous knowledge.
- **Algorithmic Bias and Misrepresentation:** AI models are a reflection of the data they are trained on. Given that this data is overwhelmingly dominated by English and Western cultural contexts, AI systems are rife with biases that misrepresent or stereotype marginalized communities. This can manifest as "cultural flattening," where AI generates inaccurate and homogenizing depictions of diverse cultures, for example, by conflating distinct Indigenous artistic styles or cultural attire. Research also shows that for low-resource languages, large language models are significantly more likely to generate harmful, toxic, or simply irrelevant responses, posing a direct safety risk.
- **The Low-Resource "Curse":** The very term "low-resource" frames the problem from a technological perspective, obscuring the fact that these languages are often rich in cultural and linguistic complexity. The technical reality, however, is that the vast majority of the world's languages lack the digital data to be visible to AI systems. Furthermore, the "curse of multilinguality" demonstrates a troubling phenomenon where, as large multilingual models are expanded to include more languages, their performance on the lowest-resource languages can actually degrade, as the model's capacity is stretched thin.

This analysis reveals a central paradox: the technologies that hold the promise of revitalizing endangered languages are often built on a data-extractive paradigm that is fundamentally at odds with the principles of Indigenous knowledge protection and self-determination. Resolving this paradox is the core challenge for this committee. The most critical innovations required are not just technical, but also social, legal, and ethical. They lie in the development of new governance models that can ensure technology is developed in service of, and under the control of, the communities it is meant to benefit.

V. "Nothing About Us Without Us": The Centrality of Indigenous Data Sovereignty

The principle of "Nothing About Us Without Us" is a global call for the full and effective participation of marginalized groups in the decisions that affect their lives. In the context of AI and language revitalization, this principle finds its most concrete expression in the movement for Indigenous Data Sovereignty (IDSov). IDSov is not an obstacle to technological innovation; rather, it is the essential framework for ensuring that innovation is ethical, sustainable, and genuinely empowering. It shifts the paradigm from one where Indigenous communities are passive subjects of technology to one where they are the sovereign authors of their own digital futures.

A. Defining Indigenous Data Sovereignty (IDSov) and Governance

Indigenous Data Sovereignty is the inherent right of Indigenous Peoples to govern the collection, ownership, and application of their own data. This right is not granted by states or corporations; it is derived from the fundamental right to self-determination as affirmed in international law. IDSov encompasses all forms of data, including:

- **Data about Indigenous peoples and their lands:** This includes demographic information, health statistics, and environmental data collected by governments and other external entities.
- **Data from Indigenous peoples:** This includes their cultural heritage, traditional knowledge, oral histories, and linguistic data.

Indigenous Data Governance is the practical exercise of sovereignty. It refers to the policies, protocols, legal agreements, and community-led institutions that Indigenous peoples create to manage their data in accordance with their own values, laws, and worldviews. This is the crucial mechanism for translating the principle of sovereignty into action.

B. Core Principles for Ethical Engagement

To guide the implementation of IDSov, Indigenous scholars and communities around the world have developed clear, actionable principles. These frameworks provide a roadmap for researchers, technology developers, and policymakers seeking to engage in ethical partnerships.

- **The OCAP® Principles (Canada):** Developed by First Nations in Canada, OCAP® is one of the most well-established frameworks for Indigenous data governance. It asserts that First Nations have the right to:
    - **Ownership:** First Nations collectively own their cultural knowledge, data, and information.
    - **Control:** First Nations communities and their representative bodies must control all aspects of data management that affect them.
    - **Access:** First Nations must be able to access data about themselves, regardless of where it is held.
    - **Possession:** First Nations have a right to the physical possession of their data. These principles provide a clear assertion of rights that must be respected in any AI project involving First Nations data.
- **The CARE Principles for Indigenous Data Governance (Global):** Developed by the Global Indigenous Data Alliance (GIDA), the CARE Principles are designed to complement the widely used FAIR data principles (Findable, Accessible, Interoperable, Reusable). While FAIR principles focus on the characteristics of data that enable machine-readability, CARE focuses on the people and purpose-oriented dimensions necessary for ethical data use. The principles are:
    - Collective Benefit: Data ecosystems must be designed and function in ways that enable Indigenous peoples to derive benefit.
    - Authority to Control: Indigenous peoples' rights and interests in their data must be recognized, and their authority to control that data must be supported.
    - Responsibility: Those working with Indigenous data have a responsibility to foster relationships with Indigenous peoples that are grounded in respect and reciprocity.
    - Ethics: Indigenous peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle. The CARE principles provide a crucial ethical layer to data management, ensuring that the focus remains on human rights and collective benefit.
    -

C. Frameworks for Ethical Partnership and Implementation

Adhering to these principles requires a fundamental shift in how technology projects are conceived and executed. The most successful and ethical initiatives are those that embody a commitment to community leadership and genuine partnership from the outset.

- **Community-Led Governance:** The principle of self-determination dictates that revitalization projects must be community-led. This means Indigenous communities must be in control of decision-making at every stage, from project conception and design to data collection, model training, and deployment. External experts, such as linguists or AI researchers, should act as collaborators and facilitators who respond to community-defined needs, rather than as directors imposing an external agenda. The ultimate

goal should be capacity building, empowering communities with the skills and infrastructure to sustain their language work independently.

- **Co-Design and Reciprocity:** Effective partnerships are built on long-term relationships, trust, and mutual respect. This requires ongoing, two-way consultation where community members are treated as equal partners and experts in their own right. It also involves reciprocity, ensuring that the project provides tangible benefits to the community, as defined by the community itself.
- **Culturally Specific Licensing and Legal Agreements:** Standard intellectual property frameworks like copyright and open-source licenses are often insufficient to protect Indigenous data, which is typically held collectively and in perpetuity. This has led to the development of innovative, culturally grounded legal tools. The
- *kaitiakitanga* (guardianship) license created by Te Hiku Media is a pioneering example, establishing a legal framework that ensures Māori data is used for the benefit of the Māori people while prohibiting commercial exploitation or harmful applications. Similarly, the Esethu Framework in Africa provides a community-centric data license that grants research access while allowing African-owned entities to use datasets commercially.
- **National Policy Alignment:** Governments have a critical role to play in creating a policy environment that recognizes and supports IDSov. New Zealand's guidance for businesses on using Māori data in AI is a leading example, explicitly incorporating principles of Māori Data Sovereignty and providing practical advice on ethical engagement. In Australia, the National Indigenous Australians Agency (NIAA) has published an AI Transparency Statement committing to culturally safe and responsible engagement with AI. These national-level policies are essential for holding both public and private sector actors accountable.

Ultimately, the evidence from successful projects and the clear articulation of principles from Indigenous communities worldwide lead to an inescapable conclusion: Indigenous Data Sovereignty is not a barrier to be overcome, but rather the essential pathway to achieving responsible, human-centric, and trustworthy AI in the context of language revitalization. For GPAI, promoting IDSov is a direct and effective means of fulfilling its core mandate.

VI. The Path Forward: Questions for Deliberation

The challenge before this committee is to translate the principles and insights discussed in this guide into a concrete, actionable framework for international cooperation. The following questions are intended to guide your deliberations and assist in the formulation of substantive draft resolutions. They are organized around the key themes of governance, funding, and technology, reflecting the multifaceted nature of the issue.

Governance and Ethics

1. How can GPAI, in alignment with its core mandate, develop and promote a global framework for the ethical development and deployment of AI for endangered languages? How can the CARE Principles for

Indigenous Data Governance (Collective Benefit, Authority to Control, Responsibility, Ethics) be integrated as the foundational pillars of such a framework?

2. What specific actions can GPAI's Data Governance Working Group take to support Indigenous communities? Could this include creating a repository of model data-licensing agreements, such as the *kaitiakitanga* license, that protect Indigenous cultural and intellectual property while enabling community-led innovation?
3. How can GPAI encourage its member states to formally align their national AI strategies with the principles of the UN Declaration on the Rights of Indigenous Peoples (UNDRIP)? What mechanisms can be proposed to ensure that Free, Prior and Informed Consent (FPIC) becomes a mandatory prerequisite for any use of Indigenous data in AI systems, whether by public or private actors?
4. What role can GPAI play in advocating for the reform of international intellectual property law to better recognize and protect the collective, intergenerational nature of traditional knowledge and oral traditions in the digital age?

Funding and Capacity Building

1. What innovative funding mechanisms can GPAI recommend or help establish to provide direct financial support for Indigenous-led AI research, the development of community-owned digital infrastructure, and language revitalization projects?
2. How can GPAI facilitate meaningful and equitable partnerships between Indigenous communities, academic institutions, and technology companies? What best practices can be established to ensure these partnerships prioritize capacity building, skills transfer, and long-term sustainability, empowering communities to become creators and governors of AI technology?
3. What recommendations can this committee make to ensure that international development aid and national funding for AI innovation include specific allocations and criteria for projects that support linguistic diversity and are governed by Indigenous communities?

Technology and Data

1. How can GPAI leverage its network of experts to support and accelerate research into AI techniques that are better suited for low-resource contexts, such as Small Language Models (SLMs), few-shot learning, and methods that reduce the need for massive datasets?
2. What concrete steps can GPAI take to advance the proposal for a "Multicultural AI Consortium" that emerged from the Tokyo Innovation Workshop? How can such a consortium be structured to ensure it is governed by Indigenous and community representatives and dedicated to the creation of high-quality, ethically sourced, and community-controlled datasets?
3. What technical standards and auditing procedures should GPAI advocate for to identify, measure, and mitigate the risks of algorithmic bias, cultural misrepresentation, and "cultural flattening" in AI models that are trained on or generate content in Indigenous languages?

4. How can GPAI promote the development and adoption of open-source tools and platforms for language revitalization, ensuring they are accessible, adaptable, and designed to support the principles of data sovereignty?