

# The Algorithmic Erasure of Linguistic Diversity: A Forensic Analysis of Artificial Intelligence, Quebec French, and the Push for Sovereign Models

The advent of Large Language Models (LLMs) represents a profound paradigm shift not merely in computational capability, but in the structural evolution of global human communication. As artificial intelligence rapidly solidifies its position as the primary interface for information retrieval, civic participation, and creative expression, the underlying architectural framework of these systems is silently dictating the boundaries of acceptable language. Because these frontier models are overwhelmingly engineered within the Silicon Valley ecosystem and trained on English-centric, Anglo-American datasets, their outputs exhibit a massive homogenization effect. This algorithmic standardization actively erodes regional dialects, cultural idioms, and the structural nuances of non-dominant languages. Ultimately, this phenomenon poses a severe, existential threat to global linguistic diversity, acting as a technological catalyst for digital language extinction.

This exhaustive research report provides a comprehensive, forensic investigation into the technical, cultural, and sociopolitical mechanisms driving AI-induced linguistic homogenization. By utilizing the unique linguistic reality of Quebec French (*le français québécois*) as a primary case study, the analysis illustrates exactly how standardized AI outputs systematically strip away the cultural soul of regional dialects, replacing them with "flattened," anglicized, or Parisian-standardized variants. Furthermore, the report examines the global threat of "Data Deserts" on minority African and Indigenous languages, and evaluates the emerging defensive paradigms. These defenses include the massive capitalization of "Sovereign AI" infrastructure and the implementation of robust linguistic public policies by the Government of Quebec and the broader *Francophonie*. The findings synthesized herein are structured as a bilingual article architecture, providing the foundational research necessary for a comprehensive investigative publication.

## 1. The Algorithmic Homogenization of Language

The dilution of linguistic diversity by artificial intelligence is not merely an accidental byproduct of user preference or an easily patched software bug; it is a profound structural artifact embedded within every stage of the LLM development pipeline. From data curation and tokenization to model alignment and generation, the algorithmic homogenization of language occurs through several compounding technical and sociolinguistic mechanisms that systematically privilege prestige languages over regional variants.

## The Overwhelming Dominance of English in Training Corpora

The foundational issue underlying AI language dilution is the massive, uncorrected over-representation of English in pre-training datasets. The internet, which serves as the primary scraping ground for frontier generative models, is disproportionately Anglophone. Consequently, the vast majority of the world's machine learning models are predominantly trained using English data drawn from Western sources.<sup>1</sup> For the 1.52 billion people who speak English, these systems operate with near-perfect fluency; however, they rely on binary internet data that inherently widens the digital language divide for the rest of the world.<sup>2</sup>

When multilingual models process queries in non-dominant languages, they do not inherently "think" in those languages. Instead, they often rely on an implicit translation mechanism or a "pivot language" strategy.<sup>4</sup> Recent interpretability studies analyzing the hidden states across the intermediate neural layers of language models reveal that multilingual processing is highly influenced by the composition of the pre-training data.<sup>4</sup> Languages heavily represented in the data, primarily English, act as dominant semantic anchors.<sup>4</sup> When a user inputs a prompt in a regional language, the model frequently maps the semantic meaning into its English-dominant latent space, generates a response conceptually rooted in English logic, and translates the output back into the target language. This process systematically strips away the cultural context, yielding an output that is structurally and philosophically Anglo-American, merely wearing a foreign linguistic mask.<sup>6</sup>

## Tokenization Costs and Structural Asymmetries

The bias against non-standard languages is further entrenched at the very first step of data processing: tokenization. Tokenizers, algorithms that break text down into machine-readable numerical chunks, are optimized for the languages and dialects most prevalent in their training data. Extensive analyses demonstrate that non-standard forms, including regional dialects, incur significantly higher segmentation costs.<sup>8</sup> Because the tokenizer does not recognize regional vocabulary as whole words, it fragments them into smaller, disjointed byte-pair tokens.<sup>8</sup>

This creates a dual penalty for linguistic diversity. First, it degrades the model's contextual understanding of the dialect, as the word loses its semantic cohesiveness. Second, it increases the computational cost—and by extension, the financial cost via API pricing—of processing non-standard text. This architectural reality creates an inherent systemic friction against linguistic diversity, making it computationally expensive for AI companies to support anything other than standard, prestige language forms.<sup>8</sup>

## The Alignment Tax and Cultural Flattening

Beyond translation mechanics, AI systems are actively standardizing how people reason and communicate, leading to what researchers at the University of Southern California term "cultural homogenization".<sup>1</sup> Because models undergo safety and quality alignment using techniques like Reinforcement Learning from Human Feedback (RLHF), they are trained to

produce outputs that human raters deem helpful, polite, and universally acceptable.<sup>10</sup> However, these raters are often trained to apply standardized, neutral criteria that penalize colloquialisms, regional slang, or culturally specific viewpoints that deviate from the mainstream consensus.<sup>1</sup>

This alignment process generates a self-reinforcing loop of "flattened" language. AI models prioritize efficiency, generalization, and broad consensus.<sup>10</sup> When human users interact with these models, they face a documented trade-off between "fidelity" (the precision with which the AI reflects the user's specific intent) and "communication cost" (the time and effort required to revise the prompt).<sup>11</sup> Because the majority of users prioritize productivity, they accept the initial, generalized outputs generated by the AI.<sup>11</sup> As this highly generic, AI-generated content floods the internet, it becomes the training data for the next generation of models. Over time, this recursive feedback loop continuously narrows the spectrum of human expression into a homogenized, "average" vernacular, stripping language of its power, emotion, and regional spice.<sup>10</sup>

<b>Mechanism of Homogenization</b>	<b>Technical Description</b>	<b>Sociolinguistic Impact</b>
<b>Data Imbalance &amp; Scraping</b>	Over 90% of frontier model training data is derived from English/Western sources, creating a dominant latent space. <sup>1</sup>	Non-English outputs reflect Anglo-American cultural norms, eroding local epistemologies and perspectives. <sup>6</sup>
<b>Implicit Pivot Translation</b>	Multilingual LLMs route complex reasoning through a dominant language (English) in intermediate neural layers. <sup>4</sup>	Loss of culture-specific idioms and untranslatable concepts; outputs sound alienating or technically unnatural. <sup>6</sup>
<b>Tokenization Friction</b>	Non-standard dialects and minority languages are heavily fragmented during tokenization due to lack of representation. <sup>8</sup>	Increased computational cost and reduced contextual understanding for regional dialects, penalizing diverse inputs. <sup>8</sup>
<b>Alignment Standardization</b>	RLHF prioritizes safe, universal, "prestige language" outputs over localized, historically rich	Subordination of dialects; users are forced to code-switch to standard languages to be

	dialectal variations. <sup>9</sup>	understood, reinforcing the bias. <sup>9</sup>
--	------------------------------------	--

## 2. The Quebec French Case Study

The threat of linguistic homogenization is acutely visible in the context of Quebec French (*le français québécois*). Quebec French is not merely an accented subset of Metropolitan (Parisian) French; it possesses a distinct vocabulary, unique syntactic structures, historical idioms, and a rich lexicon of expressions deeply tied to the province's history, harsh environment, and social evolution. When processed by models engineered in Silicon Valley and trained primarily on European French data, the Quebecois cultural essence is systematically marginalized, resulting in a phenomenon linguists and sociologists describe as "linguistic colonization".<sup>13</sup>

### Neoliberal Worldviews and the Erasure of Cultural Anchors

Technology is never neutral; it inherently carries the worldview of its creators. Research highlighted by Radio-Canada indicates that contemporary generative AI develops a worldview heavily rooted in a "neoliberal, technophile, very masculine, and American-influenced" context.<sup>13</sup> This worldview tends to flatten diverse cultural perspectives that do not align with its dominant influence. Standard AI models predominantly output a version of French that adheres strictly to Parisian standardization, frequently categorizing Quebecois regionalisms as errors, typos, or marginal data anomalies.<sup>13</sup>

A primary example of this cultural loss is the inability of frontier models to comprehend or appropriately generate *les sacres* (Quebecois swear words traditionally rooted in the terminology of the Catholic Church). These expressions are far more than mere profanity; they are complex linguistic markers that reflect the historical evolution, rebellion, and secularization of Quebec society during the Quiet Revolution.<sup>13</sup> By filtering these terms out through safety guidelines, or failing to recognize them entirely, AI models actively suppress a vital window into Quebec's shared cultural history.<sup>13</sup> Because users realize the AI cannot understand their natural dialect, they modify their own linguistic behavior, adopting standard Metropolitan French simply to be understood by the machine. This forced code-switching inadvertently reinforces the model's prioritization of the standard register, accelerating the demise of the local dialect.<sup>9</sup>

### Empirical Evidence: The QFrCoRE and QFrCoRT Benchmarks

The failure of LLMs to process the Quebec dialect has been empirically quantified through the recent introduction of two specialized benchmark datasets developed by the Group for Research in Artificial Intelligence of Laval University (GRAIL): the Quebec-French Corpus of Regional Expressions (QFrCoRE) and the Quebec-French Corpus of Regional Terms (QFrCoRT).<sup>14</sup>

The methodology for constructing these corpora involved manually extracting idioms from

specialized Quebecois dictionaries and online lexical resources.<sup>15</sup> To rigorously test the models, researchers used GPT-4o-mini to generate semantically plausible but incorrect distractors, measuring similarity using BLEU, ROUGE, and BERTScore to ensure the AI could not simply guess based on context clues.<sup>15</sup>

- **QFrCoRE** comprises 4,633 instances of idiomatic multi-word phrases, such as "*attache ta tuque avec de la broche*" (literally: "fasten your toque with wire," idiomatically meaning to brace oneself for what is coming).<sup>17</sup>
- **QFrCoRT** contains 171 regional idiomatic words, such as "*Tiguidou!*" (an expression meaning everything is great).<sup>17</sup>

When researchers evaluated 111 distinct LLMs against these benchmarks in a zero-shot setting, the results revealed a critical, systemic dialect gap.<sup>17</sup> While the models performed adequately on standard Metropolitan French, a staggering 65.77% of the models performed significantly worse on Quebec idioms.<sup>18</sup> Only 9.0% of the tested models exhibited any proficiency favoring the regional dialect.<sup>18</sup> The disparity is not due to syntactic complexity, as models failed equally on single words and full phrases (an average score difference of only 5.5% between the two corpora); rather, the failure stems entirely from a lack of regional and cultural vocabulary in the training data.<sup>17</sup> While a highly select cluster of 27 massive frontier models (such as GPT-5.1, Gemini-2.5-pro, and DeepSeek-reasoner) achieved above 80% accuracy, the vast majority of accessible, smaller open-source models experienced severe socio-linguistic dissonance.<sup>17</sup> Radio-Canada corroborated these findings, noting that roughly 40% of existing language models simply do not understand *expressions québécoises* at all.<sup>13</sup>

## Frontier Models: ChatGPT, Claude, and Mistranslations

When users prompt frontier models like ChatGPT and Claude with highly specific Quebecois contexts or invented idioms, the homogenization effect becomes starkly visible. Experiments testing ChatGPT, Gemini, and Claude on regional or fabricated idioms reveal disparate but flawed approaches: some models hallucinate entirely false histories for the idioms, while others over-explain the concepts using highly unnatural, academic phrasing rather than conversational dialect.<sup>19</sup>

While Claude AI is noted for having a sophisticated attention mechanism that can grasp standard French linguistics and syntax quite well, it frequently defaults to "prestige" responses.<sup>9</sup> If a user asks Claude to translate an emotionally charged, highly informal Quebecois sentence into English, and then asks it to translate it back, the model rarely returns the sentence to its original Quebecois form; instead, it outputs a sanitized, Parisian equivalent. This demonstrates that the model's internal representation of the language lacks the localized depth required to maintain dialectal continuity, treating the regional variant as an input to be corrected rather than a valid output to be generated.

## The Proliferation of Anglicized Syntax (*Calques*)

Compounding the erasure of regional vocabulary is the phenomenon of anglicized syntax, where AI outputs French vocabulary forcibly mapped onto English grammatical structures—commonly known in linguistic circles as *calques de l'anglais* or syntactic anglicisms.<sup>21</sup> Because LLMs frequently utilize English as an implicit pivot language during cross-lingual generation, their French outputs structurally mirror English sentence construction.<sup>4</sup>

In the Quebec context, where the preservation of the French language against the overwhelming surrounding Anglophone influence is a daily existential, political, and cultural struggle, AI-generated *calques* actively accelerate linguistic degradation. For example, AI models frequently generate phrases like "*prendre un cours*" (a direct translation of the English "to take a course") instead of the correct French "*suivre un cours*".<sup>22</sup> Similarly, "open line" radio shows are translated literally as "*ligne ouverte*" rather than "*tribune téléphonique*".<sup>22</sup>

Furthermore, sociolinguistic studies investigating the changing use of anglicisms in Quebec French highlight the increasing prevalence of morphologically unintegrated English verbs inserted directly into French syntax.<sup>23</sup> Observations include phrases such as "*Je vous rassure, on ne voulait pas sneak une proposition*" or "*Peux-tu bring le trailer quand tu vas venir?*".<sup>23</sup> Linguistic awareness campaigns point to a generational decay timeline: from "*J'ai appelé ma mère*" (2000), to "*J'ai callé ma mère*" (2010), to the fully unintegrated "*J'ai call my mom*" (2030).<sup>23</sup> As AI-generated content normalizes these anglicized structures in chatbots, automated emails, and digital marketing, the technology inadvertently acts as a Trojan horse. It bypasses simple vocabulary controls and directly attacks the structural and syntactic integrity of the French language, leaving native speakers with sentences that are "so full of English words that you can't even understand what the sentence means".<sup>24</sup>

English Source Concept	Proper French (Quebec/Standard)	AI-Generated Anglicism (Calque)	Syntactic and Cultural Impact
"To take a course"	<i>Suivre un cours</i>	<i>Prendre un cours</i>	Direct semantic mapping overriding French verb conventions, degrading grammatical standards. <sup>22</sup>
"Open line" (Radio)	<i>Tribune téléphonique</i>	<i>Ligne ouverte</i>	Literal translation resulting in a loss of contextual meaning

			and precise vocabulary. <sup>22</sup>
"To abuse" (a person)	<i>Maltraiter</i>	<i>Abuser</i> (un enfant)	Severe semantic drift; <i>abuser</i> in French means to deceive or overuse, not to inflict physical harm. <sup>21</sup>
"Fasten your seatbelt" (Metaphorical)	<i>Attache ta tuque avec de la broche</i>	<i>Attache ta ceinture</i>	Total erasure of the Quebecois regional idiom in favor of a literal, Parisian standard equivalent. <sup>17</sup>
"To call my mom"	<i>J'ai appelé ma mère</i>	<i>J'ai call my mom</i>	Morphological non-integration; English verbs forcibly inserted into French syntax without conjugation. <sup>23</sup>

### 3. The Global Threat to Linguistic Diversity

The digital crisis facing Quebec French is merely a microcosm of a much broader, catastrophic threat to global linguistic diversity. Of the approximately 7,000 languages currently spoken worldwide, an estimated 40% to 43% are classified as endangered or at imminent risk of extinction, with a minority language dying out every two weeks.<sup>25</sup> Artificial intelligence, in its current developmental trajectory, is aggressively accelerating this timeline by creating immense digital barriers for non-dominant languages, a phenomenon fundamentally driven by the existence of "Data Voids."

#### The Concept of Data Voids and Data Deserts

Data Voids, or Data Deserts, refer to digital ecosystems or market populations where limited, fragmented, or entirely non-existent digitized textual resources prevent AI systems from learning a language, producing accurate predictions, or understanding a specific community.<sup>27</sup> A stark, highly exclusionary digital divide separates the languages AI can "see" and the hundreds of languages it remains blind to.<sup>28</sup>

The statistics are sobering. While the ChatGPTs and Geminis of the world operate effectively

for the 1.52 billion people who speak English, they severely underperform for the world's 97 million Vietnamese speakers, and completely fail for the 1.5 million people who speak the Uto-Aztecan language Nahuatl.<sup>2</sup> Even within languages that boast massive populations, the lack of digitized data creates artificial deserts. For instance, Swahili has 200 million speakers but lacks sufficient digitized, computational resources for AI models to learn from, whereas a language like Welsh, despite having vastly fewer speakers, benefits from extensive documentation and digital preservation efforts, allowing it to perform better in LLM environments.<sup>2</sup>

The situation is exceptionally dire on the African continent, which boasts over 2,000 languages, constituting nearly one-third of the world's linguistic diversity. Yet, a stunning 88% of African languages are considered "severely underrepresented" or "completely ignored" in computational linguistics.<sup>25</sup> The newly released SAHARA benchmark, which rigorously evaluated 517 African languages across various AI tasks, demonstrated that languages spoken by tens of millions—such as Wolof, Hausa, Oromo, Fulfulde, and Kinyarwanda—consistently cluster among the lowest performers across reasoning, generation, and classification metrics.<sup>28</sup> This discrepancy is not due to linguistic complexity, but rather decades of systemic underinvestment in datasets and digital infrastructure.<sup>28</sup> Furthermore, deep digital disparities exist even between African languages; models trained on Afrikaans achieve strong machine translation performance due to abundant digital resources, while indigenous languages like isiZulu and Sepedi lag significantly in accuracy.<sup>29</sup>

## **Epistemic Erasure and Digital Extinction**

When a language exists within a Data Desert, the consequences extend far beyond mere user inconvenience; it results in "epistemic erasure"—the systemic deletion of a culture's specific way of knowing, categorizing, and interacting with the world.<sup>30</sup> To younger generations growing up natively interacting with smartphones, digital assistants, and AI-driven platforms, the inability of their devices to comprehend their mother tongue sends a powerful, destructive psychological message: their language is obsolete, irrelevant, and does not belong in the modern world.<sup>31</sup>

This tragedy is poignantly illustrated by the fate of the Western Shoshone dialect in North America. When elder Mae Timbimboo Parry passed away, she took with her thousands of words describing precise angles of sunlight and animal tracks. When community members attempt to use modern AI transcription tools to record the traditional ecological knowledge of surviving elders, the algorithms return error messages, failing to recognize the audio as human language at all.<sup>31</sup> Similarly, the Ambo language in Australia (holding 60,000 years of medicinal plant knowledge) and the Beeke language in Nigeria exist in a total digital void, completely invisible to modern voice-to-text algorithms.<sup>31</sup>

When AI models lack sufficient training data for a minority language, they do not degrade gracefully. They hallucinate wildly, amplify historical stereotypes, and fail to distinguish between highly divergent cultural contexts.<sup>28</sup> Attempts by tech companies to bypass the data desert

using automated translation often propagate errors. As Stanford researchers note, automated translation may produce a sentence that makes sense word-by-word, but remains "culturally completely incorrect," failing to capture the nuance of how the language is actually spoken.<sup>32</sup> Consequently, relying purely on globalized LLMs for minority languages results in severe socio-economic exclusion, effectively locking these populations out of the civic, educational, healthcare, and economic opportunities generated by the AI revolution.<sup>2</sup>

<b>Global Data Desert Demographics</b>	<b>Linguistic Status &amp; AI Capability</b>	<b>Epistemic Consequence</b>
<b>Swahili vs. Welsh</b>	Swahili (200M speakers) fails due to lack of digitized corpus; Welsh (far fewer speakers) succeeds due to digital preservation. <sup>2</sup>	Highlights that AI exclusion is driven by data availability, not population size or linguistic complexity. <sup>2</sup>
<b>SAHARA Benchmark (Africa)</b>	517 African languages tested; major languages like Wolof, Hausa, and Oromo consistently fail in reasoning tasks. <sup>28</sup>	88% of African languages remain ignored in computational linguistics, locking millions out of digital economies. <sup>25</sup>
<b>Western Shoshone (US)</b>	AI transcription tools completely fail to recognize the audio as human language, returning error messages. <sup>31</sup>	Loss of hyper-specific ecological knowledge (e.g., animal tracks, sunlight angles) that cannot be translated. <sup>31</sup>
<b>Nahuatl (Uto-Aztecan)</b>	Major frontier models completely fail to generate or reason in the language despite 1.5 million speakers. <sup>2</sup>	Forces younger generations to abandon the language in favor of Spanish/English to access digital tools. <sup>31</sup>

## **4. The Defense Strategy: Sovereign AI & Policy Frameworks**

Recognizing that linguistic sovereignty is now deeply intertwined with digital and infrastructural sovereignty, governments, linguistic organizations, and academic institutions are mounting aggressive defense strategies. The push back against algorithmic homogenization is manifesting through two primary vectors: the massive capitalization of localized "Sovereign AI"

physical infrastructure, and the implementation of aggressive legislative shields and linguistic public policies.

## The Rise of Sovereign AI in Canada and Quebec

Sovereign AI operates on the foundational premise that nations must develop and control artificial intelligence systems aligned with their own local values, trained on regional data corpuses, and hosted on domestic physical infrastructure, rather than relying entirely on foreign, Silicon Valley-based hyperscalers.<sup>34</sup> By 2026, global spending on sovereign AI systems is projected to surpass \$100 billion, driven by the need for strategic control and data residency.<sup>34</sup>

The Government of Canada has launched the *Canadian Sovereign AI Compute Strategy*, backed by a historic \$2 billion investment outlined in Budget 2024 and Budget 2025, aimed at building advanced AI supercomputing capacity domestically.<sup>35</sup> Within this framework, the federal Ministry of Artificial Intelligence and Digital Innovation initiated a call for proposals to develop large-scale AI data centers with capacities exceeding 100 megawatts.<sup>37</sup> These sovereign facilities prioritize indigenous participation, strict data residency, low-carbon energy usage, and economic ecosystem benefits.<sup>37</sup>

The province of Quebec has positioned itself as the vanguard of this movement, aggressively leveraging its vast hydroelectric resources, cool climate, and pre-existing AI research ecosystem. The provincial government recently granted \$36 million to Mila (the Quebec Artificial Intelligence Institute) to strengthen its network of excellence and accelerate the ethical development of AI aligned with Quebec's strategic and linguistic interests.<sup>39</sup> Furthermore, Mila has entered into a strategic partnership with 5C and Hypertec to develop a massive \$250 million Sovereign AI Research Hub at Hypertec's global headquarters in LaSalle, Quebec.<sup>40</sup> This next-generation campus is explicitly designed to provide up to 3 MW of secure, state-of-the-art compute capacity (utilizing NVIDIA, AMD, and Intel GPUs) cooled by advanced immersion and applied heat recovery technologies.<sup>40</sup>

This sovereign infrastructure allows Quebec researchers to train and run inference on localized models without surrendering proprietary or culturally sensitive information to foreign clouds.<sup>40</sup> Organizations like Scale AI are concurrently funding projects within the *Zone économique métropolitaine* (ZEM) in the Greater Quebec City region, entrusting funds to ensure that homegrown AI solutions prioritize local economic and social realities, thereby cementing digital sovereignty.<sup>42</sup>

By utilizing Sovereign AI infrastructure, developers can intentionally curate training datasets that ingest Quebec literature, local journalism, parliamentary records, and archives from cultural organizations like ADISQ (Association québécoise de l'industrie du disque, du spectacle et de la vidéo). Furthermore, through parameter-efficient fine-tuning (PEFT) and low-rank adaptation (LoRA), researchers can continuously pre-train models on regional dialects under tight compute budgets, mathematically forcing the model to map its neural pathways

according to Quebecois syntax, vocabulary, and cultural idioms.<sup>43</sup>

## Legislative Shields: Bill 96 and the *Francophonie*

Technological infrastructure is being simultaneously augmented by robust, unyielding policy frameworks. In Quebec, the modernization of the *Charter of the French Language* (commonly known as Bill 96 or the *Act respecting French, the official and common language of Québec*) has established stringent legal requirements designed to insulate the language from digital and corporate erosion.<sup>45</sup> The legislation firmly affirms French not just as the official language, but as the *common language* of the Quebec nation, imposing a strict "duty of exemplarity" on the civil administration to use French exclusively in public services, with highly specific, narrow exceptions for Indigenous communities and recent immigrants.<sup>47</sup>

Crucially, Quebec is introducing legislation demanding the "discoverability of and access to original French-language cultural content in the digital environment" by modifying the Charter of Human Rights and Freedoms.<sup>48</sup> This sweeping mandate requires that digital platforms, televisions, and connected devices must ensure their default interfaces and content delivery algorithms prioritize French.<sup>48</sup> For AI agents, LLM-powered enterprise tools, and automated chatbots deployed by businesses operating in Quebec, this represents a hard compliance mandate: models cannot simply translate English thoughts into broken, anglicized French; they must fundamentally operate in high-quality, culturally accurate French to avoid severe regulatory penalties and loss of market access.<sup>45</sup>

Institutions like the *Office québécois de la langue française* (OQLF) are also actively stepping in to standardize the vocabulary of the AI revolution itself to prevent the wholesale importation of English tech-jargon. By officially defining and promoting terms like *algorithme de recommandation* (recommendation algorithm), *voix* (voice memo), and *logiciel d'intelligence artificielle* (AI software), and by providing comprehensive lexicons through the *Vitrine linguistique*, the OQLF ensures that the discourse surrounding AI remains deeply anchored in the French language.<sup>49</sup> The Conseil de l'innovation du Québec further supports this by driving collective reflection on the societal impacts of AI, specifically assessing its effects on the growth and development of the national language and culture through comprehensive reports like *Prêt pour l'IA*.<sup>52</sup>

On the global diplomatic stage, the *Organisation internationale de la Francophonie* (OIF) has elevated AI linguistic diversity to a primary pillar of its mandate. Representing 93 states and governments (comprising 300 million speakers across five continents), the OIF is lobbying aggressively at international forums, including the United Nations, to ensure that linguistic diversity is codified in international AI governance treaties.<sup>55</sup> When tech executives boast about adding automated translations of 110 languages to their platforms, digital policy advisors for La Francophonie push back against these PR maneuvers.<sup>57</sup> They argue that automated translation is insufficient to prevent epistemic erasure; therefore, deep, structural linguistic diversity and localized data training must become the non-negotiable "backbone of digital policy".<sup>30</sup>

<b>Defense Vector</b>	<b>Action / Initiative</b>	<b>Strategic Objective</b>
<b>Sovereign Infrastructure</b>	\$250M Hypertec/Mila Sovereign AI Hub in LaSalle, Quebec. <sup>40</sup>	Provide domestic compute capacity to securely train models on local data without relying on foreign tech giants. <sup>40</sup>
<b>Federal Compute Strategy</b>	\$2 Billion Canadian Sovereign AI Compute Strategy. <sup>35</sup>	Build national supercomputing infrastructure to ensure long-term global competitiveness and data sovereignty. <sup>36</sup>
<b>Legislative Mandates</b>	Bill 96 & Discoverability Act (modifying the Charter of Human Rights). <sup>46</sup>	Force digital platforms and enterprise AI tools to operate in high-quality French by default, enforcing a duty of exemplarity. <sup>47</sup>
<b>Terminological Control</b>	OQLF definitions for AI vocabulary (e.g., <i>algorithme de recommandation</i> ). <sup>50</sup>	Prevent the lexical invasion of English tech-jargon and maintain the structural integrity of the French language. <sup>51</sup>
<b>Global Diplomacy</b>	OIF resolutions at the United Nations demanding structural digital diversity. <sup>56</sup>	Combat digital imperialism and ensure non-dominant languages are embedded in international AI governance frameworks. <sup>57</sup>

## 5. Contrasting Analytical Perspectives

The debate over how to manage AI's profound impact on linguistic diversity reveals deep philosophical, economic, and technical tensions within the global artificial intelligence community.

### Global Efficiency vs. Cultural Erasure

From an engineering, corporate, and economic standpoint, the centralization of AI models

around a dominant language (English) is highly efficient. Creating a single, massive monolithic model that utilizes English as a universal semantic pivot allows hyperscale companies to deploy their technologies globally at unprecedented speeds. It radically lowers the barrier to entry for developing new features, simplifies the complex alignment process (RLHF), and avoids the exorbitant, often prohibitive costs associated with scraping, curating, and cleaning massive datasets for thousands of low-resource languages. Proponents of this centralized approach argue that using automated "translation as a bridge" actually democratizes access to knowledge.<sup>32</sup> If an AI can instantly translate the sum total of human medical, scientific, and mathematical knowledge from English into a minority language, it empowers that community immediately, even if the cultural nuance of the translation is admittedly imperfect.<sup>32</sup>

Conversely, sociolinguists, ethicists, and cultural advocates view this relentless drive for efficiency as a form of algorithmic imperialism. They argue that language is not merely a neutral conduit for data transmission; it is a profound cognitive framework that fundamentally shapes a culture's worldview. When an AI processes an indigenous, African, or regional Quebecois concept through the philosophical and linguistic lens of Silicon Valley before translating it back, it alters the core meaning of the thought.<sup>6</sup> Efficiency, in this view, is achieved through the violent erasure of local epistemology, ensuring that minority languages survive only as hollowed-out, syntactic translations of Anglo-American ideals.<sup>30</sup>

## **The Paradox of Sovereign Isolation vs. Global Open-Source Collaboration**

The defense strategy of Sovereign AI also presents a highly complex technical paradox. Nations investing heavily in sovereign data centers, low-carbon power grids, and localized models (such as France, Canada, and the province of Quebec) seek to protect their sensitive data, culture, and economic independence from foreign monopolization.<sup>34</sup>

However, the development of true, frontier-level AI requires data and collaboration on a planetary scale. As highlighted by AI policy analysts, "Sovereignty in AI faces a deeper paradox, one that makes global collaboration essential, not optional".<sup>41</sup> Genuine autonomy in AI requires active participation in global open ecosystems. For example, the 2024 Nobel Prize-winning AlphaFold system, which revolutionized drug discovery, was only made possible through a global open-access molecular database.<sup>41</sup> No single nation could have assembled that dataset in isolation.

Developing a language model entirely isolated on Quebecois or French data might perfectly preserve the dialect, but it severely risks producing a model that is intellectually stunted, lacking the vast scientific, coding, and mathematical reasoning capabilities embedded in broader, global datasets.<sup>41</sup> Therefore, experts suggest the most viable path forward lies in a "third path": leveraging massive open-source foundational models (such as LLaMA) and utilizing localized Sovereign compute capacity to fine-tune these models using Parameter-Efficient Fine-Tuning (PEFT) on high-quality regional corpuses.<sup>41</sup> This hybrid approach allows a region to download global intelligence, refine it securely behind local firewalls using cultural data, and run

inference locally, successfully balancing global computational power with local cultural fidelity.<sup>41</sup>

## 6. Bilingual Key Terms (Glossary)

To navigate the complex intersection of artificial intelligence engineering, computational linguistics, and legislative preservation, a standardized vocabulary is essential. The following glossary provides English and French equivalents of highly technical and conceptual terms relevant to this domain, reflecting the standardized terminology officially endorsed by institutions such as the OQLF.<sup>50</sup>

English Term	French Equivalent	Context / Definition
<b>Algorithmic Homogenization</b>	<i>Homogénéisation algorithmique</i>	The process by which AI models reduce the variance in human communication, standardizing output to reflect a dominant (usually Anglo-American) culture. <sup>1</sup>
<b>Sovereign AI</b>	<i>IA souveraine</i>	AI physical infrastructure (data centers, GPUs) and models developed, hosted, and controlled within a specific nation or region to protect local data, security, and cultural values. <sup>34</sup>
<b>Data Desert / Data Void</b>	<i>Désert de données / Vide de données</i>	Digital markets or linguistic populations with insufficient or fragmented digitized text to train accurate, culturally aware AI models. <sup>27</sup>
<b>Parameter-Efficient Fine-Tuning (PEFT)</b>	<i>Microréglage efficace en paramètres</i>	A computational method used to adapt large pre-trained models to low-resource regional dialects (like Quebec French) updating only ~1%

		of parameters, saving compute costs. <sup>43</sup>
<b>Implicit Translation Mechanism</b>	<i>Mécanisme de traduction implicite</i>	The algorithmic phenomenon where an LLM processes non-English inputs by internally mapping and reasoning through an English-dominant latent space. <sup>4</sup>
<b>Syntactic Anglicism (Calque)</b>	<i>Anglicisme syntaxique (Calque)</i>	The direct, literal translation of an English phrase into French that violates traditional French grammatical structures (e.g., <i>prendre un cours</i> ). <sup>21</sup>
<b>Recommendation Algorithm</b>	<i>Algorithme de recommandation</i>	A predictive algorithm that provides personalized results; heavily regulated under Quebec's new digital discoverability laws. <sup>51</sup>
<b>AI Alignment</b>	<i>Alignement de l'IA</i>	The iterative process of ensuring an AI model's behavior matches human intentions, safety filters, and ethical guidelines; often the source of dialectal flattening. <sup>50</sup>
<b>Epistemic Erasure</b>	<i>Effacement épistémique</i>	The total loss of a culture's specific way of knowing, conceptualizing, and describing the world due to chronic digital and algorithmic exclusion. <sup>30</sup>
<b>Voice Memo</b>	<i>Voxto</i>	An OQLF-coined term referring to a short voice recording sent via mobile

		device, standardizing digital communication vocabulary. <sup>51</sup>
--	--	---

## 7. Bilingual Article Architecture

The exhaustive research compiled in this report serves as the architectural foundation for a compelling, long-form investigative article. To ensure the final piece resonates seamlessly with both English and French readerships—while satisfying the demands of technology analysts and cultural preservationists alike—the following narrative hook and structural outline are proposed.

### Narrative Hook / Accroche Narrative

**(English):** "In the digital age, a language dies twice. First, when the last elder stops speaking it. Second, when the algorithms refuse to recognize it. As artificial intelligence models built in Silicon Valley rapidly become the ultimate gatekeepers of global knowledge and communication, they are silently standardizing the way humanity thinks, writes, and creates. For Quebec, a province defined by its centuries-long, existential struggle for linguistic survival, the primary threat is no longer physical assimilation by a surrounding Anglophone nation, but invisible erasure by an algorithm. When ChatGPT cannot understand a historical *sacre*, or when an enterprise AI outputs French structured entirely by English grammar, it is not merely a software glitch—it is the algorithmic colonization of a culture. The fight for the future of the French language has moved from the streets to the server farms."

**(Français):** "À l'ère numérique, une langue meurt deux fois. D'abord, lorsque le dernier aîné cesse de la parler. Ensuite, lorsque les algorithmes refusent catégoriquement de la reconnaître. Alors que les modèles d'intelligence artificielle conçus dans la Silicon Valley deviennent rapidement les gardiens ultimes des connaissances et des communications mondiales, ils standardisent silencieusement la façon dont l'humanité pense, écrit et crée. Pour le Québec, une province définie par sa lutte existentielle et séculaire pour sa survie linguistique, la menace principale n'est plus l'assimilation physique par une nation anglophone environnante, mais l'effacement invisible par un algorithme. Quand ChatGPT ne comprend pas un *sacre* historique, ou lorsqu'une IA d'entreprise génère un français structuré entièrement par une grammaire anglaise, ce n'est pas un simple bogue informatique : c'est la colonisation algorithmique d'une culture. Le combat pour l'avenir de la langue française s'est déplacé des rues vers les centres de données."

### Article Outline / Plan de l'Article

#### 1. The Invisible Erasure (L'effacement invisible):

- Open with the phenomenon of "flattened" AI text and the concept of linguistic colonization. Use the empirical failure of 65% of LLMs on the QFrCoRE benchmark (e.g., the inability to comprehend "*attache ta tuque*") as the inciting incident to demonstrate the massive gap between global technology and local reality.

2. **Inside the Black Box (Dans la boîte noire):**
  - Transition deep into the technical mechanics. Explain the architectural realities of the "Data Desert," tokenization friction, and how the "English pivot bias" forces non-English thoughts through a neoliberal, Anglo-American filter. Provide concrete examples of how Claude and ChatGPT hallucinate or over-explain idioms, resulting in *calques de l'anglais* that degrade the syntactic integrity of French.
3. **The Global Fallout (Les répercussions mondiales):**
  - Broaden the scope to show that Quebec is not fighting alone. Discuss the SAHARA benchmark and the immediate existential threat AI poses to African and Indigenous languages (such as the Western Shoshone and Ambo). Frame the issue as one of global epistemic injustice and digital extinction.
4. **Building the Digital Fortress (Bâtir la forteresse numérique):**
  - Detail the fierce resistance. Highlight the massive \$250M Hypertec/Mila Sovereign AI campus in LaSalle, the Federal \$2 Billion Sovereign Compute Strategy, and the legislative shield provided by Bill 96 and the OQLF's terminological controls. Frame Sovereign AI not as political isolationism, but as a mandatory infrastructure for cultural survival.
5. **Conclusion: The Future of Thought (L'avenir de la pensée):**
  - Conclude with the philosophical and geopolitical stakes. Reiterate that preserving linguistic diversity in artificial intelligence is not merely about translating words correctly for a user interface; it is about ensuring that the future of human-machine interaction reflects the true, vibrant multiplicity of the human experience. Algorithms must be forced to learn to speak our languages, lest humanity is forced to think exclusively in theirs.

## Works cited

1. AI is changing more than your writing — it may be shaping your worldview - USC Dornsife, accessed on April 17, 2026, <https://dornsife.usc.edu/news/stories/ai-may-promote-cultural-homogenization/>
2. How AI is leaving non-English speakers behind | Stanford Report, accessed on April 17, 2026, <https://news.stanford.edu/stories/2025/05/digital-divide-ai-llms-exclusion-non-english-speakers-research>
3. How language gaps constrain generative AI development - Brookings Institution, accessed on April 17, 2026, <https://www.brookings.edu/articles/how-language-gaps-constrain-generative-ai-development/>
4. Language Dominance in Multilingual Large Language Models - ACL Anthology, accessed on April 17, 2026, <https://aclanthology.org/2025.blackboxnlp-1.7.pdf>
5. PLUG: Leveraging Pivot Language in Cross-Lingual Instruction Tuning - ACL Anthology, accessed on April 17, 2026, <https://aclanthology.org/2024.acl-long.379.pdf>
6. Do Large Language Models Have an English Accent? Evaluating and Improving

- the Naturalness of Multilingual LLMs - Apple Machine Learning Research, accessed on April 17, 2026, <https://machinelearning.apple.com/research/english-accent>
7. Discrimination by LLMs: Cross-lingual Bias Assessment and Mitigation in Decision-Making and Summarisation - arXiv, accessed on April 17, 2026, <https://arxiv.org/html/2509.09735v1>
  8. Which English Do LLMs Prefer? Quantifying American and British English Through a Postcolonial Lens | OpenReview, accessed on April 17, 2026, <https://openreview.net/forum?id=cbh3tMZHdx>
  9. Sociolinguistic Dynamics on Digital Platforms - Emergent Mind, accessed on April 17, 2026, <https://www.emergentmind.com/topics/sociolinguistic-dynamics-of-digital-platforms>
  10. A.I. Is Homogenizing Our Thoughts : r/technology - Reddit, accessed on April 17, 2026, [https://www.reddit.com/r/technology/comments/1lkzzez/ai\\_is\\_homogenizing\\_our\\_thoughts/](https://www.reddit.com/r/technology/comments/1lkzzez/ai_is_homogenizing_our_thoughts/)
  11. AI from AI: a Future of Generic and Biased Online Content? - UCLA Anderson Review, accessed on April 17, 2026, <https://anderson-review.ucla.edu/ai-from-ai-a-future-of-generic-and-biased-online-content/>
  12. The Homogenization of Language: How AI Could Flatten Our Words and Thoughts, accessed on April 17, 2026, <https://haixun.medium.com/the-homogenization-of-language-how-ai-could-flatten-our-words-and-thoughts-c6526804c746>
  13. Protéger la diversité des langues face à l'IA | Radio-Canada, accessed on April 17, 2026, <https://ici.radio-canada.ca/nouvelle/2246151/diversite-langue-quebecoise--ia-intelligence-artificielle>
  14. Paper page - A Set of Quebec-French Corpus of Regional Expressions and Terms, accessed on April 17, 2026, <https://huggingface.co/papers/2510.05026>
  15. [Literature Review] A Set of Quebec-French Corpus of Regional Expressions and Terms, accessed on April 17, 2026, <https://www.themoonlight.io/en/review/a-set-of-quebec-french-corpus-of-regional-expressions-and-terms>
  16. Quebec-French Idiom Benchmark Datasets | PDF - Scribd, accessed on April 17, 2026, <https://www.scribd.com/document/929479411/2510-05026v1>
  17. A Set of Quebec-French Corpus of Regional Expressions and ... - arXiv, accessed on April 17, 2026, <https://arxiv.org/pdf/2510.05026>
  18. [2510.05026] Idiom Understanding as a Tool to Measure the Dialect Gap - arXiv, accessed on April 17, 2026, <https://arxiv.org/abs/2510.05026>
  19. I invented a fake idiom to test AI chatbots — only one called my bluff | Tom's Guide, accessed on April 17, 2026, <https://www.tomsguide.com/ai/i-invented-a-fake-idiom-to-test-ai-chatbots-only-one-called-my-bluff>

20. "Wait this is fucking insane - Claude immediately guessed I was French" :  
r/ClaudeAI, accessed on April 17, 2026,  
[https://www.reddit.com/r/ClaudeAI/comments/1hs7t3p/wait\\_this\\_is\\_fucking\\_insane\\_claude\\_immediately/](https://www.reddit.com/r/ClaudeAI/comments/1hs7t3p/wait_this_is_fucking_insane_claude_immediately/)
21. Dites-le en français | United Nations, accessed on April 17, 2026,  
<https://www.un.org/en/node/67818>
22. Anglicismes : Éviter les emprunts inutiles | PDF | Langue française - Scribd,  
accessed on April 17, 2026,  
<https://fr.scribd.com/document/395800780/Anglicismes>
23. Investigating attitudes towards a changing use of anglicisms in Quebec French |  
Canadian Journal of Linguistics/Revue canadienne de linguistique, accessed on  
April 17, 2026,  
<https://www.cambridge.org/core/journals/canadian-journal-of-linguistics-revue-canadienne-de-linguistique/article/investigating-attitudes-towards-a-changing-use-of-anglicisms-in-quebec-french/1B1BA00CA929A3F32C4235E848008223>
24. Investigating attitudes towards a changing use of anglicisms in Quebec French,  
accessed on April 17, 2026, <https://muse.jhu.edu/article/961875>
25. How the African Languages Lab empowers low-resource languages - Smartling,  
accessed on April 17, 2026,  
<https://www.smartling.com/blog/african-languages-lab-empowers-low-resource-languages-with-ai>
26. Embracing AI to preserve dying languages - FairPlanet, accessed on April 17,  
2026,  
<https://www.fairplanet.org/story/embracing-artificial-intelligence-to-preserve-dying-languages/>
27. How algorithmic data deserts exclude consumers - MIT Sloan, accessed on April  
17, 2026,  
<https://mitsloan.mit.edu/ideas-made-to-matter/how-algorithmic-data-deserts-exclude-consumers>
28. When AI Can't Understand Your Language, Democracy Breaks Down |  
TechPolicy.Press, accessed on April 17, 2026,  
[https://www.techpolicy.press/when-ai-cant-understand-your-language-democracy-breaks-down-/](https://www.techpolicy.press/when-ai-cant-understand-your-language-democracy-breaks-down/)
29. Bridging the AI Divide: DSFSI Drives Multilingual Knowledge Access at UP's  
'Abstracts into Indigenous Voices' Event | University Of Pretoria, accessed on April  
17, 2026,  
<https://www.up.ac.za/afriidsai/news/bridging-ai-divide-dsfsi-drives-multilingual-knowledge-access-ups-abstracts-indigenous-voices>
30. From Paris to Prompt: How La Francophonie Quietly Changed AI's Language  
Priorities, accessed on April 17, 2026,  
<https://multilingual.com/la-francophonie-ai-language-diversity/>
31. The AI Bias Nobody Talks About: How Training Data Is Erasing Minority Languages,  
accessed on April 17, 2026,  
<https://medium.com/@afolabidare50/the-ai-bias-nobody-talks-about-how-training-data-is-erasing-minority-languages-65c81707c9fa>

32. Closing the Digital Divide in AI | Stanford HAI, accessed on April 17, 2026, <https://hai.stanford.edu/news/closing-the-digital-divide-in-ai>
33. Linguistic diversity in AI and ML: Why it's important, accessed on April 17, 2026, <https://sigma.ai/linguistic-diversity-in-ai/>
34. Sovereign AI: Why Nations are Treating Compute as Critical Infrastructure. - RAISE Summit, accessed on April 17, 2026, <https://www.raisesummit.com/post/sovereign-ai-compute-critical-infrastructure>
35. Canadian Sovereign AI Compute Strategy - Innovation, Science and Economic Development Canada, accessed on April 17, 2026, <https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy>
36. Canada launches national initiative to build large-scale AI supercomputing capacity, accessed on April 17, 2026, <https://www.canada.ca/en/innovation-science-economic-development/news/2026/04/canada-launches-national-initiative-to-build-large-scale-ai-supercomputing-capacity.html>
37. Enabling large-scale sovereign AI data centres, accessed on April 17, 2026, <https://ised-isde.canada.ca/site/ised/en/enabling-large-scale-sovereign-ai-data-centres>
38. Government of Canada launches call for proposals for large scale sovereign AI data centres, accessed on April 17, 2026, <https://www.dlapiper.com/en/insights/publications/2026/02/government-of-canada-launches-call-for-proposals-for-large-scale-sovereign-ai-data-centres>
39. The Quebec Government Grants \$36M to Mila to Strengthen Artificial Intelligence Research and Talent, accessed on April 17, 2026, <https://mila.quebec/en/news/the-quebec-government-grants-36m-to-mila-to-strengthen-artificial-intelligence-research-and>
40. Mila, 5C and Hypertec Announce \$250 Million LaSalle Campus and Sovereign AI Research Hub to Strengthen Canada's Role as a Global Leader in AI Innovation, accessed on April 17, 2026, <https://mila.quebec/en/news/mila-5c-and-hypertec-announce-250-million-lasalle-campus-and-sovereign-ai-research-hub-to>
41. Sovereignty Is Not Solitude: Open Source as Canada's Third Path in AI, accessed on April 17, 2026, <https://www.cigionline.org/articles/sovereignty-is-not-solitude-open-source-as-canadas-third-path-in-ai/>
42. Québec to support innovative AI projects across the Zone économique métropolitaine (ZEM), accessed on April 17, 2026, <https://www.scaleai.ca/quebec-to-support-innovative-ai-projects-in-the-zone-economique-metropolitaine-zem/>
43. Low-Resource Dialect Adaptation of Large Language Models: A French Dialect Case-Study - arXiv, accessed on April 17, 2026, <https://arxiv.org/html/2510.22747v3>
44. [2510.22747] Low-Resource Dialect Adaptation of Large Language Models: A French Dialect Case-Study - arXiv, accessed on April 17, 2026, <https://arxiv.org/abs/2510.22747>
45. Bill 96 Quebec Language Law Impact on Businesses - Smartcat, accessed on April

- 17, 2026, <https://www.smartcat.com/blog/bill-96/>
46. Charter of the French Language - Wikipedia, accessed on April 17, 2026, [https://en.wikipedia.org/wiki/Charter\\_of\\_the\\_French\\_Language](https://en.wikipedia.org/wiki/Charter_of_the_French_Language)
47. Modernization of the Charter of the French language ..., accessed on April 17, 2026, <https://www.quebec.ca/en/government/policies-orientations/french-language/modernization-charter-french-language>
48. Québec introduces Bill requiring discoverability of French-language content on digital platforms - Dentons Data, accessed on April 17, 2026, <https://www.dentonsdata.com/quebec-introduces-bill-requiring-discoverability-of-french-language-content-on-digital-platforms/>
49. OQLF - Vocabulaire de l'intelligence artificielle - Gouvernement du Québec, accessed on April 17, 2026, <https://www.oqlf.gouv.qc.ca/vocabulaire-intelligence-artificielle>
50. vocabulaire intelligence artificielle.pdf - Symboles, accessed on April 17, 2026, <https://www.oqlf.gouv.qc.ca/ressources/bibliotheque/dictionnaires/VocabulairesPDF/vocabulaire-intelligence-artificielle.pdf>
51. Les 12 termes de l'année 2025 de l'Office québécois de la langue ..., accessed on April 17, 2026, [https://www.oqlf.gouv.qc.ca/office/communiqués/2025/20251201\\_douze\\_termes\\_annee\\_2025.aspx](https://www.oqlf.gouv.qc.ca/office/communiqués/2025/20251201_douze_termes_annee_2025.aspx)
52. SYNTHÈSE DE LA JOURNÉE D'ÉTUDE COMMUNS DU, accessed on April 17, 2026, <https://espace.inrs.ca/id/eprint/16327/1/SyntheseCommunsIA.pdf>
53. Dans l'œil de l'Obvia : notes de breffage pour les acteurs publics, accessed on April 17, 2026, <https://www.obvia.ca/dans-loeil-de-lobvia>
54. Intelligence artificielle | Conseil de l'innovation du Québec, accessed on April 17, 2026, <https://conseilinnovation.quebec/intelligence-artificielle/>
55. IP24083 | Geopolitics of the French Language - RSIS, accessed on April 17, 2026, <https://rsis.edu.sg/rsis-publication/idss/ip24083-geopolitics-of-the-french-language/>
56. Cooperation between the UN and the OIF is based on common objectives - France ONU, accessed on April 17, 2026, <https://onu.delegfrance.org/cooperation-between-the-un-and-the-Oif-is-based-on-common-objectives>
57. Mind your language: The battle for linguistic diversity in AI - UN News, accessed on April 17, 2026, <https://news.un.org/en/story/2025/03/1161406>