

Artificial Intelligence and the Possibilities of Linguicide: Legitimate Fears About African Indigenous Languages

Philip Chika Omenukwa, Judith Mambu Nghibi

Catholic Institute of West Africa (CIWA)

DOI: <https://doi.org/10.47772/IJRISS.2026.100600943>

Received: 22 June 2026; Accepted: 27 June 2026; Published: 09 July 2026

ABSTRACT

While modern breakthroughs in artificial intelligence (AI) offer unprecedented computational utility, their global proliferation mounts severe pressure on marginalized linguistic heritages. This study investigates the impact of AI and natural language processing (NLP) systems on African indigenous languages, examining the threat of linguicide—the systematic decimation of a language that occasions mental and social death. Employing a critical phenomenological approach, this paper analyzes how the architectural optimization of AI influences social relations, cultural identity, and human agency. We demonstrate that AI training datasets inherently privilege high-resource world languages, reinforcing historical colonial hierarchies and perpetuating epistemic injustice and epistemicide. Furthermore, the study traces how the computational drive for standardization revives the logical positivist illusion of a “perfect language,” flattening localized semantic nuances and accelerating linguistic assimilation among youth demographics. Ultimately, this paper argues that linguicide is a structural vulnerability portended by AI centralizing tendencies. To counter this digital marginalization, we propose a philosophical framework of linguistic pluralism grounded in decolonial computing, participatory data curation, and structural policy reform, asserting that the defense of indigenous mother tongues is a radical act of epistemic sovereignty.

Keywords: Artificial Intelligence, Linguicide, Epistemicide, Epistemic Injustice, Indigenous Languages, Cultural Identity, Worldhood, Linguistic Homogenization

INTRODUCTION

Language is not merely an external tool for communication; it is the primary medium through which human beings construct and make sense of reality (Emmitt & Komesaroff, 2003). It encodes a community’s worldview, ethical values, and internal structures of meaning-making. As Hans-Georg Gadamer (2004) argues, language forms the “horizon” within which understanding is possible; it fundamentally shapes how we interpret ourselves, others, and the world. Language, in this sense, is not secondary to thought, but constitutive of it.

This view resonates strongly with Ludwig Wittgenstein’s (1953) later philosophy. In his *Philosophical Investigations*, Wittgenstein rejects the early positivist ideal of a logically perfect language, positing instead that the meaning of a word is determined by its dynamic use within localized language-games—rule-governed activities rooted in shared “forms of life.” Every language, therefore, is an expression of the existential way of life of its speakers, embodying unique modes of acting, perceiving, and valuing (Okonkwo, 2024; Omenukwa, 2024a). This paradigm shift in Wittgenstein’s philosophy affirms that meaning arises not from abstract, universal structures, but from practical usage within a distinct cultural context (Omenukwa, 2024a).

When a language disappears, the unique worldview it encodes—its metaphors, epistemic categories, values, and localized modes of reasoning—vanish with it. This is not simply an isolated linguistic loss, but a profound erosion of human diversity and cultural identity. It represents a concrete case of linguicide (Fanon, 1963; Ngũgĩ, 1986). This paper investigates how contemporary artificial intelligence (AI) systems exacerbate this threat, analyzes the philosophical challenges this digital assimilation poses, and proposes frameworks for linguistic justice.

Language and Epistemic Injustice

The systematic marginalization and gradual decimation of African indigenous languages demonstrate a gross display of epistemic injustice, a concept theorized by Miranda Fricker (2007). Epistemic injustice occurs when individuals or communities are wronged specifically in their capacity as knowers. In the context of language, this injustice manifests when certain linguistic forms are systematically excluded from education, governance, or digital technologies, thereby silencing the knowledge systems embedded within them.

This purposeful, obliterating dynamic of linguistic dominance is captured by Ngũgĩ wa Thiong'o (1986) through the concept of *linguicide*. Linguicide is demonstrable when a dominant culture imposes its own linguistic categories as the universal standard of assessment, rationality, and digital belonging, thereby assuming a supervisory role over the autonomous development of another natural language. The dominant linguistic culture carves out narratives for the marginalized group, forcing it to adopt structural habits completely alien to its natural configurations.

In *The Wretched of the Earth*, Frantz Fanon (1963) highlights the violence of historical narratives and the structural injury they inflict on the social psyche. This aligns with Chinua Achebe's (1999) assertion that European colonialism inflicted severe trauma on the African psyche—a trauma that modern AI algorithms risk replicating. Fanon observed that narratives function as instruments of systemic violence when used for the mental and theoretical oppression of a people, leading to a loss of existential meaning and an ontological schizophrenia.

Linguicide occurs when a dominant technological or cultural paradigm systematically erodes the values of another culture through the subtle obliteration of its linguistic meanings and symbols, consequently destroying self-identity and historical consciousness. Because cultural meaning and symbolism are embedded within language, once a language loses its social value, its speakers are pressured to assume the cultural values of the dominant language. Fanon (1963) illustrates the profound psychological, social, cultural, political, and economic implications of such institutionalized oppression, making a case for structural decolonization and a return to localized linguistic analysis.

Traditional ecological knowledge, indigenous moral codes, and community-based epistemologies are often transmitted orally through native languages. When these languages are delegitimized by algorithmic exclusion, they are rendered epistemically invisible, endangering their continued survival. AI systems, trained predominantly on a narrow subset of global languages, contribute directly to this injustice by codifying only a small fragment of global knowledge. As Omenukwa and Kanu (2024a) argue, indigenous languages carry the “worldhood” of their speakers—their distinctive experience of being-in-the-world. To erase these languages is to disrupt alternative ontologies and epistemologies, perpetuating what Gayatri Spivak (1988) defines as the epistemicide of historically marginalized peoples.

Language as a Site of Power and Resistance

Language is a terrain where power is continuously enacted, maintained, and contested. Michel Foucault (1972) described discourse as a primary apparatus for producing knowledge and exercising power, showing that dominant languages are never neutral, but function as instruments of structural control. The global hegemony of English in science, technology, and international commerce reflects entrenched colonial and neo-colonial power structures, a reality accelerated by modern globalization (Omenukwa, 2025a). Globalization breaks formerly restricted boundaries, opening nations to unprecedented contact, domination, and cultural friction, which frequently results in an inevitable implosion of foreign linguistic registers into indigenous constitutions (Omenukwa, 2025a).

This dominance is reinforced by artificial intelligence. Natural language processing (NLP) models, including large language models (LLMs) like ChatGPT, Google Translate, and virtual assistants, overwhelmingly privilege English and a few global high-resource languages. This side-lining occurs not because indigenous languages lack intrinsic value, but because they lack digitized data corpora. This data-driven exclusion reproduces a strict linguistic hierarchy in which dominant voices are amplified while alternative voices are systematically muted.

However, indigenous and marginalized languages remain critical sites of resistance. As postcolonial thinkers demonstrate, the conscious act of speaking, writing, and thinking in one's native tongue is a profound political gesture—a refusal to be entirely assimilated into a uniform digital empire (Fanon, 1963; Ngũgĩ, 1986). Every natural language constitutes a complete universe of encounter and exploration deserving equal philosophical, cultural, and technological legitimacy (Omenukwa, 2024a). Therefore, the defense of indigenous languages against the linguistic centralization of AI is an essential act of epistemic and cultural sovereignty.

Underrepresentation in AI Systems as a Structural Threat

Data Bias and the Logic of Optimization

Despite the existence of roughly 7,000 spoken languages globally, fewer than 200 are represented in modern NLP systems, and fewer than 50 are robustly supported by advanced architectures like OpenAI's GPT series or Google Translate. This disparity stems from an uneven distribution of digitized language data, which privileges high-resource languages with massive online footprints and formal institutional backing.

The situation is acute for African indigenous languages due to historical systemic neglect, colonial language policies, and inadequate funding for linguistic digitization. This underrepresentation excludes speakers from participating fully in digital spaces and threatens to erase cultural identities intrinsically bound to language.

AI systems depend fundamentally on massive datasets to train neural networks (Hay et al., 1960; Kanu et al., 2024). The mathematical logic of optimization in machine learning prioritizes accuracy, statistical fluency, and performance metrics in widely used languages. Consequently, marginalized languages that lack extensive digitized text corpora are treated as statistical outliers or ignored entirely. This design ethos privileges scalability over cultural inclusion, reinforcing hierarchical valuations where low-resource languages are treated as computationally irrelevant. Because language and culture are strictly interdependent, the structural depreciation of the language inevitably precipitates the erosion of the culture (Omenukwa, 2024b).

The Disappearance of Oral and Unwritten Languages

The digital era presents distinct challenges for oral and unwritten languages, which account for a substantial portion of African linguistic diversity. These languages often lack standardized orthographies or formal written documentation, making them exceptionally difficult to integrate into standard, text-based machine-learning training datasets (Homa-Saldaña et al., 2025; Pinhanez et al., 2024). Crucially, many indigenous languages are preserved and transmitted through embodied performance, oral storytelling, and communal rituals—complex forms of communication that actively resist flat digital codification.

Efforts to digitally archive or algorithmically model these oral traditions introduce significant ethical complications regarding data governance:

Epistemic Agency: Who determines which linguistic elements are recorded, how they are interpreted, and the forms in which they are digitally preserved?

Decontextualization: Reducing a living, oral language to static phonetic or semantic data risks stripping it of the dynamic cultural lifeworld that sustains it.

Archival Death: When economic pressures incentivize the adoption of dominant global languages, preserving an indigenous language solely in a database without communal use results in archival death—where a language is stored technologically but ceases to be spoken, practiced, or transformed.

Thus, the AI-mediated marginalization of oral languages represents a deeper crisis of cultural dislocation and systemic epistemic invisibility.

AI and the Ideology of a “Perfect Language”: Historical Echoes From Logical Positivism to Modern NLP

The modern ambition of NLP carries deep philosophical undertones reminiscent of the historical pursuit of a universal, unambiguous, and logically flawless “perfect language.” This ideal traces back to Gottfried Wilhelm Leibniz’s *characteristica universalis*, and was later revived by 20th-century logical positivists like Bertrand Russell, Rudolf Carnap, and the Vienna Circle, who argued that language must conform to rigid empirical verification and formal logic to be meaningful (Carnap, 1963; Russell, 1903, 1912).

Wittgenstein’s (1922) early work, the *Tractatus Logico-Philosophicus*, epitomizes this tradition, asserting that language must mirror the logical structure of atomic facts. Expressions that failed to conform to this verifiable structure—such as those found in metaphysics, ethics, or indigenous cosmologies—were dismissed as nonsensical. While Wittgenstein (1953) explicitly rejected this reductionist framework in his later work, the drive for logical purification remains embedded in the computational structures of AI, which privileges syntactic consistency and semantic uniformity over ambiguity, contextual nuance, and pluralism.

Modern AI systems inadvertently revive this positivist ideology. Large language models do not employ the symbol-based rules of classical positivism; instead, they rely on probabilistic statistical models and vector embeddings. However, their ultimate operational objective is functionally identical: they optimize for standard grammatical correctness and mathematical coherence. In doing so, they reproduce historical exclusions, failing to accommodate the structural irregularities, oral registers, and localized idioms of marginalized languages.

Homogenization and Standardization in AI Communication

The operational logic of models like GPT or BERT relies on pattern recognition across massive datasets. These systems do not merely replicate human speech; they regularize it—smoothing out regional idioms, standardizing grammar, and aligning output with dominant linguistic registers. Consequently, the language generated by AI subtly enforces a regime of digital linguistic conformity (Foucault, 1972). Dialects, regional expressions, and culturally specific metaphors are filtered out or flagged as non-standard errors.

This standardization reflects a broader economic valuation of uniformity over diversity. Driven by socioeconomic incentives, speakers frequently pivot toward global languages, allowing English to act as the primary digital gatekeeper (Omenukwa, 2025a). The systemic prioritization of these standard languages in AI design optimizes computational efficiency at the direct expense of linguistic pluralism.

Is a “Perfect Language” Desirable or Oppressive?

While an optimized, highly uniform AI language interface may appear to improve global interoperability, it raises profound ethical concerns regarding cultural erasure. A monolithic linguistic framework flattens the interpretive complexities of human expression. As Paul Ricoeur (1976) notes, language is the essential condition for the possibility of interpretation and self-understanding. Imposing a highly standardized digital linguistic model constrains both communication and thought.

For marginalized communities whose languages encode distinct ontologies and moral systems, clarity enforced by external computational metrics can become a form of structural violence. Linguistic imperialism operates not only through overt coercion, but through the internal assimilation of inferiority, causing speakers to abandon their mother tongues for languages deemed “useful” or “universal” (Ngũgĩ, 1986). Every natural language constitutes an independent universe of meaning that cannot be short-changed or substituted without profound loss. In the era of AI, technical precision cannot be bought at the cost of epistemic diversity.

Youth Culture, AI, and Linguistic Assimilation: Digital Spaces and Generational Shifts

In contemporary society, youth interaction with AI occurs continuously through ubiquitous tools: virtual assistants, smart speakers, automated translation interfaces, and generative AI tutors. These interfaces are structurally non-neutral; they actively shape linguistic behavior by requiring compliance with high-resource

world languages. Whether utilizing voice interfaces or text-based generative models, young users navigate environments that privilege mainstream idioms and Western grammatical structures.

Because these digital tools are trained on dominant data corpora, indigenous languages remain functionally invisible within algorithmic spaces. This dynamic causes a severe rupture in the cultural worldhood of young speakers, who are the most likely to internalize digital linguistic norms (Omenukwa, 2024a). To achieve intelligibility and legitimacy within these algorithmic environments, younger generations rapidly adapt their speech patterns, shifting toward dominant languages for daily digital convenience and utility. This shift alters how young people conceptualize identity, community belonging, and valid knowledge systems.

Cultural Prestige and the Decline of Local Languages

AI platforms accelerate linguistic assimilation by reinforcing the cultural prestige of dominant global languages, specifically English. Within digital spaces, fluency in a dominant language is structurally linked with professional success, technical intelligence, and modernity, whereas the use of an indigenous or marginalized language can carry false connotations of low socioeconomic mobility (Okonkwo, 2024).

Pierre Bourdieu's (1991) theory of linguistic capital explains this dynamic: specific languages accrue symbolic and economic value that can be converted into social advantage. Young people gravitate toward dominant registers because those forms are structurally and economically rewarded within schools, corporate workplaces, and digital networks. This adaptation occurs at the direct expense of indigenous languages, which are mischaracterized as economic barriers rather than cultural assets. The abandonment of native tongues for digital and social acceptance disrupts historical memory, cultural continuity, and localized moral frameworks (Omenukwa & Kanu, 2024a).

The Influence of AI-Mediated Platforms

The physical and software limitations of popular AI interfaces exacerbate this linguistic decline. Virtual assistants often support fewer than 30 languages globally and prioritize accents from high-income regions. These systems frequently struggle with non-Western phonologies and indigenous syntax, communicating a clear signal to users that their natural mode of speech is invalid within the digital ecosystem.

As Wittgenstein (1922) famously observed, the limits of a speaker's language mean the limits of their world. When young speakers find their native tongues excluded from the technological systems mediating their daily lives, their expressive and interpretive capacities are artificially constrained. Unless AI architectures drastically expand their linguistic capacity and cultural inclusivity, they will continue to function as powerful engines of linguistic assimilation and cultural dislocation.

Toward a Philosophy of Linguistic Pluralism in the Age of AI

Ethical and Political Imperatives

In the context of rapid AI advancement, the preservation of linguistic diversity is an urgent ethical and political imperative. As UNESCO (2001) affirms in its Universal Declaration on Cultural Diversity, linguistic diversity is a fundamental guarantee of cultural diversity, requiring active protection in both physical and digital spheres. The systemic exclusion of marginalized languages from AI architectures infringes upon basic linguistic human rights—the right of communities to cultivate, transmit, and interact freely in their own languages (Skutnabb-Kangas, 2000).

Neglecting these rights re-enacts historical patterns of epistemic injustice by elevating dominant world languages as the sole authorized vehicles of digital knowledge. The devaluation of indigenous mother tongues deeply diminishes the cultural and existential self-understanding of their communities (Omenukwa, 2024a). Because AI serves as a primary infrastructure of contemporary knowledge production, its design must be rigorously critiqued for how it privileges specific epistemic frameworks while erasing others.

Decolonizing Language Technologies

Decolonial theory provides an essential critical framework for resisting the centralizing, homogenizing forces of AI. Coloniality—the enduring structural imbalances left by historical colonialism—persists through contemporary digital epistemologies and linguistic practices (Mignolo, 2009). Language technologies developed primarily in the Global North and modeled exclusively on Western linguistic patterns operate as instruments of digital colonialism when they marginalize local tongues.

To decolonize AI language systems, developers must challenge the implicit assumption that certain languages are inherently more valuable or “data-worthy” than others. This requires a profound rethinking of what constitutes valid linguistic data. A decolonial approach mandates participatory design—developing language technologies with and by indigenous communities rather than imposing external systems for them. No language is inherently superior or inferior; each contains an expansive universe of human thought and experience (Omenukwa & Kanu, 2024a). Furthermore, developers must resist reducing language to transactional utility, recognizing that it operates fundamentally as a carrier of culture, memory, and worldview (Emmitt & Komesaroff, 2003; Ngũgĩ, 1986).

AI Design for Linguistic Justice

Achieving linguistic justice requires concrete technological, participatory, and policy reforms:

Dimension.	Practical	Implementation Strategy
------------	-----------	-------------------------

Participatory Data Curation	Engaging native speakers, local educators, and community institutions directly to curate, validate, and govern linguistic datasets, ensuring cultural accuracy and informed consent.
-----------------------------	--

Algorithmic Inclusion	Allocating specific research and computational resources to develop low-resource NLP models, zero-shot learning architectures, and speech-to-speech models tailored for unwritten or oral languages.
-----------------------	--

Policy & Governance Mandates	Implementing state and international policy frameworks that compel technology firms to support local linguistic diversity in public-sector AI applications.
------------------------------	---

While establishing these frameworks involves technical challenges—such as a lack of standardized orthographies for certain oral languages and profit-driven corporate models—achieving linguistic justice is ultimately a question of political and moral will rather than technical capability. As Paul Ricoeur (2000) observes, justice requires a continuous, historical commitment to institutional fairness and structural inclusion. A philosophy of linguistic pluralism must unite ethical critique with pragmatic, community-led design to transform AI into a tool for localized empowerment rather than digital assimilation.

CONCLUSION

This study has evaluated the structural threats that artificial intelligence poses to marginalized and indigenous African languages. Grounded in Wittgenstein’s (1953) insights on language-games and Gadamer’s (2004) horizon of understanding, we have demonstrated how data biases and optimization algorithms reinforce historical linguistic hierarchies and perpetuate epistemic injustice. By tracing the historical parallels between the computational drive of modern NLP and the logical positivist search for a “perfect language,” this paper illustrated how statistical standardization flattens cultural nuance. Furthermore, the analysis of youth culture revealed how modern digital interfaces act as agents of linguistic assimilation, accelerating the decline of local languages due to pressures of digital utility and shifting linguistic capital.

These intersections raise a fundamental question regarding the neutrality of technology. As Martin Heidegger (1977) warned, technology is never a neutral tool; it is a structural mode of revealing that frames the world according to an internal logic of efficiency and quantification. Within AI development, this framing systematically excludes alternative knowledge systems that cannot be easily scaled or digitized. Therefore,

preserving a language within digital systems cannot be a simple act of extraction; it must be a holistic preservation of the cultural lifeworld that gives the language its vitality.

Digital inclusion must not be reduced to creating a static archive for academic curiosity. Language is a living engine of communal memory, identity, and social agency (Ngũgĩ, 1986). Consequently, technical preservation must be accompanied by active cultural revitalization, intergenerational transmission, and local community ownership of data assets.

Moving forward, interdisciplinary collaboration between philosophers, computational linguists, computer scientists, and indigenous community leaders is essential. Future research must focus on designing open-source computational architectures that empower low-resource communities to co-create their digital futures. Ultimately, the pursuit of linguistic justice within AI design is a clear moral imperative. As Paul Ricoeur (2000) reminds us, true justice lies in our capacity to listen to the voice of the Other. By designing technology that honors and reflects global linguistic pluralism, we move toward a digital humanity rooted in mutual recognition and shared epistemic sovereignty.

REFERENCES

1. Achebe, C. (1999). *Echi di ime taa bu gboo* [Odenigbo lecture]. Assumpta Press.
2. Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.). Harvard University Press.
3. Carnap, R. (1963). *Autobiography*. In P. A. Schilpp (Ed.), *The philosophy of Rudolf Carnap* (Vol. 11). Open Court.
4. Emmitt, M., & Komesaroff, L. (2003). *Language and learning: An introduction for teaching*. Oxford University Press.
5. Fanon, F. (1963). *The wretched of the earth*. Grove Press.
6. Foucault, M. (1972). *The archaeology of knowledge* (A. M. Sheridan Smith, Trans.). Pantheon Books.
7. Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
8. Gadamer, H.-G. (2004). *Truth and method* (J. Weinsheimer & D. G. Marshall, Trans.; 2nd rev. ed.). Continuum.
9. Hay, J. C., Martin, F. C., & Wightman, C. W. (1960). The mark-1 perceptron—design and performance. *Proceedings of the IRE*, 48(3), 398–403.
10. Heidegger, M. (1977). *The question concerning technology and other essays* (W. Lovitt, Trans.). Harper & Row.
11. Homa-Saldaña, C. J., Perez Perez, J. E., & Toro Galeano, M. L. (2025). Artificial intelligence in the preservation of native languages and bridging the information access gap for indigenous peoples. *Journal of Enabling Technologies*, 19(1), 63–75.
12. Kanu, A. I., Pilani, M. P., & Omenukwa, P. C. (2024). Philosophy of artificial intelligence, biblical wisdom and the exodus (31: 1-11) narrative: Converging perspectives. *Journal of Propulsion Technology*, 45(3), 3030–3044.
13. Mignolo, W. D. (2009). Epistemic disobedience, independent thought and de-colonial freedom. *Theory, Culture & Society*, 26(7–8), 159–181.
14. Ngũgĩ wa Thiong'o. (1986). *Decolonising the mind: The politics of language in African literature*. Heinemann.
15. Okonkwo, P. (2024). *The ethics of digital spaces in contemporary African discourse*. Assumpta Press.
16. Omenukwa, P. C. (2024a). On the interdependence between language and culture. *Journal of Contemporary Research in Social Sciences*, 6(2), 61–71.
17. Omenukwa, P. C. (2024b). Religion and the dynamics of language and culture in Nigeria. *West African Journal of Arts and Social Sciences*, 5(1), 142–158.
18. Omenukwa, P. C. (2025a). Language, world of existence and entrepreneurship. In *Philosophy of entrepreneurship: Intersections and insights* (pp. 112–128). CIWA Publications.
19. Omenukwa, P. C., & Kanu, A. I. (2024a). Meaning in Wittgenstein and African indigenous languages. *Pakistan Journal of Life and Social Sciences*, 22(1), 661–676.

20. Omenukwa, P. C., & Kanu, A. I. (2024b). Private language in Wittgenstein and the Igbo-African worldview. *Pakistan Journal of Social Sciences*, 22(1), 646–660.
21. Pinhanez, C., Cavalin, P., Storto, L., Finbow, T., Cobbinah, A., Nogima, J., ... & Gonçalves, I. (2024). Harnessing the power of artificial intelligence to vitalize endangered indigenous languages: Technologies and experiences. *arXiv preprint arXiv:2407.12620*.
22. Ricoeur, P. (1976). *Interpretation theory: Discourse and the surplus of meaning*. Texas Christian University Press.
23. Ricoeur, P. (2000). *The just* (D. Pellauer, Trans.). University of Chicago Press.
24. Russell, B. (1903). *The principles of mathematics*. Cambridge University Press.
25. Russell, B. (1912). *The problems of philosophy*. Williams & Norgate.
26. Skutnabb-Kangas, T. (2000). *Linguistic genocide in education—or worldwide diversity and human rights?* Lawrence Erlbaum Associates.
27. Spivak, G. C. (1988). Can the subaltern speak? In C. Nelson & L. Grossberg (Eds.), *Marxism and the interpretation of culture* (pp. 271–313). University of Illinois Press.
28. UNESCO. (2001). *Universal declaration on cultural diversity*.
29. Wittgenstein, L. (1922). *Tractatus logico-philosophicus* (C. K. Ogden, Trans.). Kegan Paul.
30. Wittgenstein, L. (1953). *Philosophical investigations* (G. E. M. Anscombe, Trans.). Blackwell.