Canadian Journal of Marketing Research
ISSN: 0829-4836
https://canadian-jmr.com/index.php/cjmr

Research Article

AI-Driven Translation, Linguistic Hegemony, and Neocolonialism: Reconsidering Language Power in the Digital Age

Prangya Priyadarsini Mohapatra[1*] and Mrunmayee Prava Pattanaik[2]
[1]Department of Humanities and Social Sciences ITER, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India
[2]Student, Faculty of Arts, Communication and Indic Studies Sri Sri University, Cuttack, Odisha, India

***Corresponding Author**
Dr. Prangya Priyadarsini Mohapatra
(prangyamohapatra@soa.ac.in)

Article History
Received: 21.07.2026
Accepted: 18.08.2026
Published: 24.09.2026

DOI: 10.61336/cjmr.1603.99

Abstract: Artificial intelligence (AI)-driven translation is frequently presented as a technology capable of reducing linguistic barriers and widening access to information. Yet translation technologies do not operate outside histories of unequal linguistic, economic, cultural, and technological power. This conceptual research paper examines the intersection of AI-driven translation, linguistic hegemony, and neocolonialism, with particular attention to data inequality, technological ownership, market concentration, cultural homogenisation, and the digital language divide. Drawing on the conceptual material supplied for this study and a critical review of scholarship on data colonialism, algorithmic bias, low-resource machine translation, and participatory language technology, the paper argues that the central problem is not AI translation itself but the unequal structures through which languages become visible, measurable, profitable, and technically supported. The paper develops a framework linking four dimensions of power: data, infrastructure, markets, and representation. It further considers how these dimensions affect marginalised and low-resource languages through uneven training data, limited evaluation resources, dependence on externally owned infrastructures, and the possible flattening of culturally embedded meanings. At the same time, the study recognises that AI can support language preservation, multilingual education, and cross-cultural communication when communities participate meaningfully in data creation, model development, evaluation, and governance. The paper concludes by proposing a community-centred and linguistically plural model of AI translation grounded in data sovereignty, participatory design, transparent evaluation, equitable access, and human oversight.

Keywords: AI-Driven Translation, Linguistic Hegemony, Neocolonialism, Data Colonialism, Low-Resource Languages, Cultural Homogenisation, Digital Language Divide, Language Technology

INTRODUCTION

Translation has historically been more than the mechanical transfer of words from one language to another. It mediates knowledge, cultural memory, political communication, education, commerce, and relations between communities. The increasing use of artificial intelligence has transformed this mediation. Neural machine translation, multilingual language models, speech technologies, and generative AI can process and produce multilingual content at a scale that conventional human translation cannot easily match. AI-driven translation therefore has genuine potential to widen

access to information, facilitate international communication, support education, and make linguistic resources available to communities that have historically faced barriers to knowledge.

However, technological efficiency should not be confused with linguistic neutrality. The systems that translate languages are trained, evaluated, deployed, and maintained within particular social and economic environments. The conceptual material underlying this paper identifies several connected concerns: dominant languages receive disproportionate representation in training data; AI development and infrastructure are concentrated among powerful corporations and countries; less widely represented languages may receive weaker technological support; and automated translation can reproduce dominant cultural assumptions or reduce culturally specific meanings. These concerns are consistent with broader scholarship showing that data and algorithmic systems can reproduce existing social inequalities rather than simply reflecting neutral technical processes [1-2].

The concept of linguistic hegemony is useful for understanding these processes. A language becomes hegemonic not merely because it has many speakers but because it acquires institutional, economic, political, educational, and technological power. When a small group of languages dominates digital infrastructures, search systems, educational resources, and AI applications, speakers of other languages may encounter a structural disadvantage. The resulting digital language divide can affect access to information, education, economic opportunity, and participation in online public life. UNESCO has similarly highlighted the importance of education in languages that learners understand and the educational consequences of linguistic exclusion [3].

The notion of neocolonialism extends this analysis from language to power. Neocolonialism refers to forms of indirect domination in which economic, political, technological, or cultural influence can persist after formal colonial rule has ended. In the digital context, the concept of data colonialism is particularly relevant. Couldry and Mejias [4] describe contemporary data extraction as a process through which human life is increasingly rendered into data and appropriated for economic and institutional purposes. Applied carefully to AI translation, this perspective raises questions about who supplies language data, who controls the infrastructures that process it, who owns resulting technologies, who benefits economically, and whose cultural assumptions become embedded in computational systems. This paper therefore approaches AI-driven translation as a sociotechnical phenomenon rather than a purely computational one. Its objective is not to portray AI translation as inherently oppressive, nor to romanticise human translation as automatically neutral. Instead, it examines the conditions under which AI translation can reproduce linguistic hierarchies and the conditions under which it can contribute to linguistic inclusion. The paper develops a critical framework around five interconnected areas: data acquisition and governance; technological infrastructure and ownership; economic and market control; cultural representation; and the digital language divide.

Research Objectives

- To examine the relationship between AI-driven translation and linguistic hegemony
- To analyse how data, infrastructure, ownership, and market concentration can reproduce unequal relations between dominant and marginalised languages
- To explore the relevance of neocolonialism and data colonialism to contemporary language technologies without treating the analogy as identical to historical colonialism
- To investigate the cultural and linguistic risks associated with automated translation, including loss of nuance, representational bias, and cultural homogenisation
- To identify community-centred and policy-oriented principles for more equitable AI-driven translation

Research Questions

- How can AI-driven translation reinforce existing forms of linguistic hegemony
- In what ways do data ownership, technological infrastructure, and market concentration shape the representation of languages in AI translation
- How might automated translation contribute to cultural homogenisation or the reduction of culturally specific meaning
- What are the implications of the digital language divide for marginalised and low-resource language communities
- What principles can guide a more equitable, culturally responsive, and community-centred model of AI-driven translation

Literature Review

Scholarship on AI and language increasingly demonstrates that computational language systems are shaped by the datasets, institutions, and values surrounding them. Bender *et al.* [1] caution against treating large language models as

context-free systems and emphasise the importance of dataset documentation, stakeholder values, and the social and environmental costs of large-scale AI. Their work is significant for translation because the apparent fluency of an output does not by itself establish cultural accuracy, representational fairness, or social legitimacy.

Noble [2] provides a related critical perspective through the concept of algorithmic oppression. Her analysis demonstrates that algorithmic systems can reproduce discriminatory social structures because search and information systems are embedded within commercial and cultural arrangements. Although her principal focus is search technology rather than machine translation, the argument is relevant to translation because both systems participate in the organisation and distribution of knowledge. The central implication is that technical mediation can privilege some representations while marginalising others.

The concept of data colonialism further expands this discussion. Couldry and Mejias [4] argue that contemporary data relations can reproduce extractive patterns historically associated with colonialism. For language technologies, this raises questions about the status of linguistic data as an economic and cultural resource. Language data are not simply technical inputs; they may contain oral histories, cultural practices, indigenous knowledge, personal communications, and community-specific expressions. Consequently, extracting such data without meaningful participation, consent, recognition, or benefit-sharing can create forms of asymmetrical value creation.

Research on low-resource machine translation provides an important counterpoint to purely critical accounts. The No Language Left Behind project explicitly addressed the challenge of extending machine translation to more than 200 languages and evaluated thousands of translation directions using human-translated benchmarks [5]. Similarly, participatory work associated with Masakhane demonstrates that communities can play an active role in building datasets, benchmarks, and translation systems for African languages [6-7]. These initiatives show that the technological landscape is not fixed: design choices, research priorities, and institutional arrangements can change the distribution of linguistic resources.

At the same time, scholarship on language technology warns that simply adding more languages does not automatically remove linguistic inequality. Joshi *et al.* [8] describe a digital language divide in which a relatively small number of languages receive disproportionate technological attention. Bird [9] further argues for greater attention to the languages and communities that have historically been underrepresented in natural language processing. Research on low-resource machine translation also emphasises that low-resource status is not merely a matter of insufficient text; it can reflect broader structural conditions involving funding, institutional capacity, community participation, and research priorities [6].

Ethical scholarship has increasingly moved toward participatory approaches. Haroutunian [10], for example, examines ethical concerns surrounding low-resource machine translation through an Armenian case study and argues for collaboration with stakeholders and attention to complementary language technologies. Mager *et al.* [11] similarly foreground the ethical concerns that arise when machine translation is developed for Indigenous languages whose linguistic forms are closely connected to cultural and community identity. These studies suggest that the question is not simply whether a language can be translated by AI, but who defines usefulness, accuracy, acceptable risk, and cultural fidelity.

Conceptual Framework: From Linguistic Hegemony to Digital Language Power

The paper proposes a four-layer framework for analysing power in AI-driven translation:

- Data power
- Infrastructural power
- Market power
- Representational power. A fifth layer—community agency—cuts across all four and provides a basis for intervention

Data Power

AI translation depends on data. Training data determine which languages, dialects, registers, genres, and cultural contexts are visible to a model. If one language is represented through extensive, diverse, high-quality corpora while another is represented through small or noisy datasets, the difference can become a difference in computational performance. The problem is therefore not only that some languages have fewer digital resources; it is that resource inequality can become technologically reproduced.

The conceptual material supplied for this study identifies data sourcing, ownership, governance, accessibility, privacy, annotation, and labour as key research areas. These dimensions are interconnected. Data may be collected from communities that have little control over subsequent use; annotation may depend on poorly compensated labour; and culturally sensitive language material may be processed without adequate privacy protections. and Mejias's [12] account of data colonialism helps frame these questions as issues of power and value rather than merely data engineering.

Infrastructural Power

AI translation also depends on computational infrastructure, cloud platforms, specialised hardware, software ecosystems, and research institutions. When these resources are concentrated in a limited number of corporations and regions, communities that depend on external systems may have limited influence over how their languages are represented. Technological dependence can therefore become a form of structural dependence even without direct political control.

This does not mean that every internationally developed AI system constitutes neocolonialism. Such a conclusion would be too broad. Rather, the neocolonial lens is analytically useful when there is a sustained asymmetry in control, resource extraction, decision-making, and benefit distribution. The distinction matters because it keeps the concept historically responsible while allowing researchers to investigate contemporary continuities in unequal technological relations.

Market Power

The translation industry is undergoing significant transformation as automated systems become integrated into localisation, customer service, education, media production, software development, and international commerce. Market concentration can influence which languages receive investment and which services are commercially prioritised. Where a language market is perceived as economically small, commercial incentives may not support the same level of model development, evaluation, or human quality control available for high-resource languages.

AI can nevertheless lower entry barriers for individuals and small organisations by providing rapid multilingual drafting and communication. The relationship is therefore ambivalent: AI translation may democratise access while simultaneously concentrating technological ownership. The key research question becomes who captures the value created by expanded multilingual communication and whether local language professionals and communities participate in that value creation.

Representational Power

Translation is interpretive because meaning is shaped by context, social relationships, cultural references, idioms, metaphors, humour, ritual language, and historical memory. Automated systems may produce fluent sentences while missing these dimensions. The resulting problem is not always an obvious grammatical error. A translation may be syntactically acceptable yet culturally reductive.

Bender *et al.* [1] and Noble [6] provide complementary reasons for treating such representational problems seriously. Large-scale language systems learn patterns from data shaped by existing social structures, while algorithmic systems can reproduce unequal visibility. In translation, this may manifest through the selection of dominant equivalents, simplification of culturally specific expressions, or the repeated use of categories that are more familiar in high-resource languages.

AI-Driven Translation and Neocolonial Structures

Neocolonialism traditionally describes forms of indirect domination in which formal political independence coexists with continued economic, political, or cultural dependence. The present study does not claim that AI translation is equivalent to colonial rule. Instead, it asks whether certain arrangements surrounding AI translation can reproduce structural patterns associated with dependency: external control over resources, dependence on foreign infrastructures, unequal economic returns, and cultural asymmetry.

The first relevant mechanism is data extraction. Language data can be treated as a resource that supports commercial and technological innovation. Yet the social origin of that data can become invisible once it enters a computational pipeline. A community may contribute linguistic material while having little knowledge of how it will be processed, reused, commercialised, or incorporated into future systems. A more equitable approach requires data governance that recognises communities as stakeholders rather than merely sources of training material.

The second mechanism is technological dependence. If translation services for a language are available primarily through external platforms, institutions may become dependent on infrastructures over which they have little control. This can create vulnerabilities concerning pricing, access, privacy, service continuity, model updates, and language policy. Local capacity-building, open research infrastructures, interoperable standards, and public-interest language technology can reduce this dependence.

The third mechanism concerns cultural authority. Translation systems can influence how texts are circulated and understood. When a dominant language becomes the main bridge through which local knowledge enters global digital spaces, the source language may become positioned as secondary. This is especially important when translation is treated as the default route to visibility rather than one component of multilingual communication. Linguistic equality should therefore not mean merely translating everything into a dominant language; it should also mean strengthening communication among and within less dominant languages.

Linguistic Hegemony in AI Translation

Linguistic hegemony operates through institutional preference. A language becomes technologically powerful when it is consistently prioritised in search, software, education, commerce, speech recognition, machine translation, and generative AI. The conceptual material for this study identifies English as a central example of a dominant language in digital environments. This does not mean that English is universally dominant in every context, but it illustrates how linguistic power can become embedded in digital infrastructures.

Language representation has at least three dimensions: quantity, quality, and diversity. Quantity concerns the amount of data available. Quality concerns the accuracy, cleanliness, and contextual richness of the data. Diversity concerns whether the data represent dialects, registers, genders, age groups, regions, social contexts, oral traditions, and culturally specific usage. A dataset may therefore be large while still producing poor representation if it excludes important varieties or relies heavily on translated material.

Recent research describes language-modeling bias as a tendency for language technologies to favour some languages, dialects, or sociolects over others in representational terms [13]. This is important because computational inclusion can be superficial. A language may technically appear on a list of supported languages while receiving substantially weaker performance or limited functionality. Meaningful inclusion therefore requires transparent, language-specific evaluation rather than simple claims of multilingual coverage.

Cultural Homogenisation and Loss of Nuance

Cultural homogenisation is one of the most important concerns raised by the conceptual framework. Language encodes relationships, values, ecological knowledge, social hierarchy, humour, emotion, ritual, and collective memory. When translation systems privilege direct lexical equivalents, culturally embedded meanings may be weakened. Idioms may be translated literally, metaphors may lose their historical associations, honorific systems may be flattened, and culturally specific terms may be replaced by approximate concepts in the target language.

The risk becomes particularly significant when automated translations are reused as source material for subsequent AI systems. A culturally simplified translation can become training data for another system, creating a feedback loop in which the reduced representation is repeatedly reproduced. Human review, community evaluation, and preservation of original-language material can interrupt such cycles.

At the same time, cultural preservation should not be understood as freezing languages in an unchanging form. Languages evolve through contact, migration, media, education, and technological use. AI translation can support this evolution by creating new opportunities for speakers to write, publish, teach, and communicate in their own languages. The critical question is whether technological change expands community agency or replaces community agency with externally defined standards.

The Digital Language Divide

The digital language divide refers to unequal technological support across languages. The supplied research material identifies unequal access to online content, software and platform support, AI translation, and digital communication as central dimensions of this divide. The consequences extend beyond convenience. When users cannot access information, educational materials, public services, or digital economic opportunities in a language they understand, linguistic inequality can become an information inequality.

UNESCO's Global Education Monitoring work emphasises that learners benefit from education in languages they understand and reports that large numbers of people lack access to education in a language they speak or understand [3]. This makes language technology particularly relevant to educational equity. If AI translation is

available primarily for already well-supported languages, it may reproduce the same inequalities that multilingual technology is expected to reduce.

However, digital language inequality is not solved by technological coverage alone. A translation tool may exist while remaining inaccessible because of cost, connectivity, interface design, privacy concerns, inadequate evaluation, or lack of community trust. Therefore, digital inclusion should be assessed through a broader framework encompassing availability, affordability, usability, accuracy, cultural adequacy, and community control.

Language Data, Labour, and Ethical Governance

Another underexamined dimension is labour. Training data often require annotation, transcription, cleaning, translation, validation, and evaluation. These tasks may be performed by workers whose contribution remains invisible in the final technology. The conceptual material for this paper specifically identifies annotation labour and inadequate compensation as areas requiring ethical attention.

A responsible translation ecosystem should recognise language work as skilled labour. Human translators, interpreters, teachers, community researchers, annotators, and language activists possess knowledge that cannot be reduced to interchangeable data points. Their participation is particularly important for low-resource and Indigenous languages, where linguistic knowledge may be concentrated among a relatively small number of speakers or community institutions.

Data governance should therefore address consent, privacy, attribution, compensation, intellectual property, benefit sharing, and the right of communities to influence how culturally sensitive materials are used. Mager *et al.* [11] demonstrate why Indigenous-language technologies require attention to community perspectives, while Haroutunian [10] similarly argues for stakeholder collaboration in low-resource machine translation.

The Transformative Potential of AI Translation

A critical analysis should not overlook the constructive possibilities of AI translation. The NLLB project demonstrates that multilingual machine translation can be expanded beyond a small set of high-resource languages, while participatory initiatives such as Masakhane demonstrate the value of community-led research and open collaboration [5-6].

AI translation can assist in language revitalisation by supporting the creation of educational materials, bilingual resources, searchable corpora, subtitles, speech interfaces, and digital archives. It can also help speakers communicate across linguistic boundaries without requiring every interaction to pass through a single dominant language. For multilingual countries such as India, this potential is especially relevant because linguistic diversity is part of everyday social, educational, and cultural life.

The goal, therefore, should not be to reject AI translation but to redesign the conditions under which it is produced and used. A language technology ecosystem can become more equitable when communities participate in defining datasets, evaluation criteria, acceptable uses, and governance structures. The shift is from a model of technology delivered to communities toward technology developed with and, where appropriate, by communities.

Implications for India and the Global South

The issues discussed here have particular relevance to India, where multilingual communication operates across regional languages, dialects, English, and multiple scripts. AI translation can create significant benefits in education, public communication, healthcare information, commerce, and cultural dissemination. Yet multilingual abundance does not automatically produce technological equality. Languages can differ substantially in the quantity and quality of digital corpora, availability of language professionals, standardisation practices, speech resources, and computational benchmarks.

A Global South perspective also requires attention to infrastructure and institutional capacity. Dependence on external cloud services, proprietary models, and commercially controlled platforms can limit local research autonomy. Public universities, regional-language institutions, libraries, archives, civil-society organisations, and local technology communities can therefore play an important role in developing open datasets, language resources, and evaluation frameworks.

The objective should be multilingual technological sovereignty rather than technological isolation. Sovereignty in this context means the capacity to participate meaningfully in decisions about language data, infrastructure, standards, and applications. International collaboration remains valuable, but collaboration should be reciprocal rather than extractive.

Toward an Equitable Model of AI-Driven Translation

The analysis suggests seven principles for developing a more equitable translation ecosystem.

- **Community Participation:** Language communities should be involved in dataset creation, annotation, model evaluation, and decisions about acceptable uses. Participation should extend beyond symbolic consultation
- **Data Sovereignty:** Communities and institutions should have meaningful rights concerning culturally sensitive language data, including consent, access, attribution, governance, and benefit sharing
- **Transparent Evaluation** Translation quality should be reported by language, dialect, domain, and task rather than through a single aggregate score. Human and community evaluation should complement automated metrics
- **Open and Distributed Infrastructure:** Open datasets, open evaluation resources, interoperable standards, and publicly accessible research infrastructures can reduce dependence on a small number of technology providers
- **Fair Language Labour:** Translation, annotation, transcription, and evaluation should be recognised as skilled work and compensated fairly, particularly where community expertise is essential
- **Cultural and Linguistic Preservation:** Systems should preserve source-language material and document culturally specific terms rather than treating them as noise to be normalised
- **Human Oversight and Accountability:** AI translation should remain subject to human review in high-stakes contexts and should provide mechanisms for reporting, correcting, and learning from harmful or culturally inappropriate outputs

DISCUSSION

The central argument of this paper is that AI-driven translation should be understood as a site of linguistic power. The technology can reduce barriers, but it can also redistribute visibility and authority unevenly. The decisive variables are not only model architecture or translation accuracy. They include who controls data, who owns infrastructure, which languages receive investment, whose labour is recognised, how cultural adequacy is evaluated, and whether communities have meaningful decision-making power.

The concept of neocolonialism is most productive when used as a structural question rather than a predetermined conclusion. The presence of a multinational company, an English-language interface, or an externally developed model does not by itself establish neocolonial domination. The concept becomes analytically stronger when researchers demonstrate concrete relationships of dependence, extraction, unequal control, and unequal distribution of benefits. This approach avoids both technological determinism and overly broad historical analogy.

Similarly, linguistic hegemony should not be reduced to the simple dominance of English. Hegemony can occur wherever one language or language variety acquires disproportionate institutional and technological authority. The relevant issue is therefore the architecture of choice: which languages are easy to search, translate, publish, teach, speak to machines, or use in professional environments, and which require users to work around technological limitations.

The paper also identifies an important tension between scalability and cultural specificity. Large-scale AI systems are often designed to serve many languages simultaneously, while culturally meaningful translation may require local knowledge and domain-specific expertise. Participatory and decentralised models offer one way to address this tension. The experience of Masakhane demonstrates that community-led research can generate resources and benchmarks while strengthening local research capacity [6-7].

Limitations and Scope for Further Research

This study is conceptual and literature-based. It does not conduct a controlled comparison of specific AI translation systems, nor does it claim to measure translation accuracy across languages. Consequently, its conclusions concern structures, risks, and governance principles rather than the performance of any particular commercial model.

Future empirical research should test the framework through comparative studies of language pairs, including English–Odia, Hindi–Odia, and other Indian-language combinations. Such studies could evaluate idioms, culturally specific expressions, gendered language, honorifics, proverbs, oral narratives, and domain-specific terminology using both automated metrics and community-based human assessment.

Further research should also investigate the economics of language data, the working conditions of annotators and translators, the representation of dialects, and the role of public institutions in building open multilingual resources. Longitudinal studies could examine whether increased AI availability strengthens language use or gradually shifts users toward dominant languages.

CONCLUSION

AI-driven translation is often described as a tool for overcoming linguistic barriers, but its social consequences depend on the structures through which the technology is developed and distributed. The analysis in this paper shows that data

inequality, technological concentration, market control, representational bias, and cultural homogenisation can interact to reproduce forms of linguistic hierarchy in digital environments.

The concept of neocolonialism provides a critical vocabulary for examining these unequal relationships, while the concept of linguistic hegemony clarifies how technological systems can elevate some languages over others. Data colonialism further draws attention to the extraction and monetisation of linguistic and social data. Taken together, these perspectives show that the politics of AI translation concerns not only whether a sentence is translated correctly, but also whose language is represented, whose knowledge is visible, whose labour is valued, and who has authority over the technological infrastructure.

Yet the future is not predetermined. Projects that expand low-resource translation and participatory language technology demonstrate that AI can also be used to strengthen linguistic inclusion. The challenge is to move from a model in which communities are merely represented by AI systems to one in which they participate in shaping those systems.

A genuinely inclusive future for AI translation should therefore be multilingual not only in output but also in governance. It should treat languages as living cultural systems rather than interchangeable datasets; recognise communities as knowledge holders rather than data sources; support local research capacity rather than deepen technological dependence; and combine computational innovation with human, cultural, and ethical judgement. Such an approach can transform AI-driven translation from a possible mechanism of linguistic hierarchy into an infrastructure for linguistic plurality, equitable knowledge access, and meaningful intercultural communication.

REFERENCES

1. Bender, E.M. *et al.* "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, Association for Computing Machinery, 2021, pp. 610–623. https://doi.org/10.1145/3442188.3445922.
2. Noble, S.U. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press, 2018. https://doi.org/10.18574/nyu/9781479833641.001.0001.
3. UNESCO. "If You Don't Understand, How Can You Learn?" *Global Education Monitoring Report*, 3 Feb. 2023. https://www.unesco.org/gem-report/en/publication/if-you-dont-understand-how-can-you-learn.
4. Couldry, N. and U.A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press, 2019. https://doi.org/10.1515/9781503609754.
5. NLLB Team *et al.* "No Language Left Behind: Scaling Human-Centered Machine Translation." *arXiv*, 2022. https://arxiv.org/abs/2207.04672.
6. Nekoto, W. *et al.* "Participatory Research for Low-Resourced Machine Translation: A Case Study in African Languages." *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp. 2144–2160. https://doi.org/10.18653/v1/2020.findings-emnlp.195.
7. Orife, I. *et al.* "Masakhane—Machine Translation for Africa." *arXiv*, 2020. https://arxiv.org/abs/2003.11529.
8. Joshi, P. *et al.* "The State and Fate of Linguistic Diversity and Inclusion in the NLP World." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, 2020, pp. 6282–6293. https://doi.org/10.18653/v1/2020.acl-main.560.
9. Bird, S. "Decolonising Speech and Language Technologies." *Proceedings of the 28th International Conference on Computational Linguistics*, International Committee on Computational Linguistics, 2020, pp. 3504–3519. https://doi.org/10.18653/v1/2020.coling-main.313.
10. Haroutunian, L. "Ethical Considerations for Low-Resourced Machine Translation." *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, Association for Computational Linguistics, 2022, pp. 44–54. https://doi.org/10.18653/v1/2022.acl-srw.5.
11. Mager, M. *et al.* "Ethical Considerations for Machine Translation of Indigenous Languages: Giving a Voice to the Speakers." *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, vol. 1, Long Papers, Association for Computational Linguistics, 2023. https://aclanthology.org/2023.acl-long.268/.
12. Couldry, N. and U.A. Mejias. "Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject." *Television & New Media*, vol. 20, no. 4, 2019, pp. 336–349. https://doi.org/10.1177/1527476418796632.
13. Joshi, P. *et al.* "Diversity and Language Technology: How Language Modeling Bias Causes Epistemic Injustice." *Ethics and Information Technology*, vol. 25, 2023, article 20. https://doi.org/10.1007/s10676-023-09742-6.