Decolonizing African NLP: A Survey on Power Dynamics and Data Colonialism in Tech Development

Nan Yan
Brooklyn College
The City University of New York
New York, USA
nan.yan80@bcmail.cuny.edu

Cheng Xu
School of Computer Science
University College Dublin
Dublin, Ireland
cheng.xu1@ucdconnect.ie

Abstract

This research paper explores the imperative to decolonize African Natural Language Processing (NLP) by addressing power dynamics and data colonialism within technology development. Based on the historical context of colonialism in Africa, the paper examines the pervasive influence of colonial traces on the research and development of NLPs, highlighting the marginalization of African languages, cultures, and voices. Through an analysis of power dynamics, the paper advocates for diversifying representation within the NLP community, empowering local communities, and challenging Eurocentric frameworks to foster more inclusive and equitable technology development. Additionally, the paper explores the concept of data colonialism and its implications for African NLP, highlighting the need for data sovereignty, community ownership, and ethical data practices. Case studies and examples illustrate the transformative potential of decolonial approaches within African NLP, while future directions outline pathways for advancing the decolonial agenda through interdisciplinary collaboration, policy advocacy, and community engagement. Ultimately, the paper calls for collective action and solidarity within the NLP community to dismantle colonial legacies and forge a more just and inclusive digital future for Africa and beyond.

1 Introduction

Natural Language Processing (NLP) stands as a cornerstone of modern technological advancement, empowering applications ranging from machine translation systems (Nekoto et al., 2020; Orife et al., 2020) to large language models (García-Díaz et al., 2023; Nguyen et al., 2023). However, beneath its promising veneer lies a complex landscape rife with power imbalances and historical injustices, particularly pronounced in the African context (Tang et al., 2023). In this paper, we embark on a critical exploration of the imperative to decolonize African NLP, shedding light on the pervasive influence of colonial legacies and data colonialism in technology development.

The discourse surrounding decolonization resonates deeply within the realm of NLP, where dominant Western perspectives, datasets, and methodologies often eclipse indigenous knowledge systems and local voices (Valiani, 2022; Redvers et al., 2020; Adams, 2021). As we navigate this terrain, it becomes increasingly evident that true progress in African NLP requires a reckoning with the power dynamics inherent in technology production. By interrogating these dynamics and advocating for more inclusive and equitable approaches, we strive to pave the way for a future where technology serves as a tool for empowerment rather than entrenchment of existing hierarchies.

Central to our inquiry is the notion of data colonialism, which underpins the extraction, exploitation, and commodification of data from African contexts. This phenomenon not only perpetuates historical injustices, but also reinforces asymmetrical power relations, leaving marginalized communities at the mercy of external interests (Couldry & Mejias, 2019; Calzati, 2021). In challenging data colonialism, we envision a paradigm shift toward data sovereignty and ethical data practices, in which

African communities reclaim agency over their digital narratives and participate meaningfully in shaping technological futures (Nhemachena et al., 2020).

Through case studies, critical analysis, and a call to action, this paper seeks to galvanize efforts towards decolonizing African NLP. By amplifying underrepresented voices, advocating for equitable access to resources, and championing alternative paradigms of technology development, we aspire to forge a path towards a more just and inclusive digital future for Africa and beyond.

2 BACKGROUND

The legacy of colonialism continues to cast a long shadow over the African continent, shaping socio-political structures, economic systems, and cultural narratives (Shizha, 2016). From the imposition of arbitrary borders to the exploitation of natural resources, colonial powers left indelible marks on the fabric of African societies, perpetuating inequalities that endure to this day (Berman et al., 2014). Within the realm of technology, this legacy manifests in myriad ways, influencing everything from access to infrastructure to the design of algorithms (Genger, 2020; Farrands, 2000).

In the context of NLP, colonial legacies are particularly pronounced, with Western epistemologies and methodologies often taking precedence over indigenous knowledge systems and local languages Xu & Yan (2023). The predominance of English, French, and other colonial languages in NLP datasets and research publications marginalizes African languages and perpetuates linguistic hegemony, reinforcing existing power differentials (Blodgett et al., 2020). Moreover, the concentration of NLP research and development in Western institutions further exacerbates these disparities, limiting opportunities for African scholars and technologists to contribute meaningfully to the field (Joshi et al., 2020).

Against this backdrop, the call for decolonization emerges as a powerful rallying cry, demanding a reevaluation of entrenched power dynamics and a commitment to equitable knowledge production. Decolonizing African NLP requires not only diversifying datasets and research methodologies but also challenging the underlying assumptions and biases that inform technology development (Falola, 2022; Schwartz, 2022). It entails centering African perspectives, languages, and priorities in the design and implementation of NLP systems, thereby fostering a more inclusive and culturally responsive approach to technological innovation.

Furthermore, the concept of data colonialism provides a critical lens through which to understand the exploitation of African data by external actors for economic gain and geopolitical control. As African nations increasingly digitize their economies and societies, they become vulnerable to data extraction and surveillance by multinational corporations and foreign governments, perpetuating patterns of dependency and exploitation reminiscent of colonial-era resource extraction (Le Billon, 2017; Corkin, 2016). Addressing data colonialism requires not only safeguarding data sovereignty and promoting ethical data practices but also challenging the asymmetrical power relations that underpin the global data economy Calzada (2021).

In the pages that follow, we delve deeper into the complexities of decolonizing African NLP, interrogating power dynamics, and advocating for alternative paradigms of technology development. By situating our analysis within the broader historical and socio-political context of colonialism in Africa, we aim to shed light on the enduring legacies of oppression and inequality that continue to shape the contours of technological innovation on the continent.

3 THE NEED FOR DECOLONIZATION IN AFRICAN NLP

Within the landscape of NLP, the urgency for decolonization in Africa is starkly evident. Despite the rich linguistic diversity and cultural heritage of the continent, African languages are frequently relegated to the periphery of NLP research and development, overshadowed by dominant Western languages (Kessi et al., 2020). This marginalization not only perpetuates linguistic hegemony but also reinforces broader patterns of coloniality, wherein Western epistemologies and methodologies dictate the direction of technological innovation.

One of the primary consequences of this imbalance is the proliferation of bias and discrimination within NLP systems. By prioritizing Western languages and cultural norms, NLP models often

fail to accurately represent the linguistic nuances and socio-cultural contexts of African languages and communities. This leads to biased outcomes, algorithmic discrimination, and exacerbation of existing inequalities, particularly for marginalized groups (Dacon, 2022). Without addressing these biases and centering African perspectives in NLP research, we risk perpetuating colonial power dynamics and entrenching patterns of exclusion and oppression (Marivate, 2020).

Moreover, the dominance of Western institutions and funding agencies in NLP research exacerbates disparities in access to resources and opportunities for African scholars and technologists (Field et al., 2021). Limited access to data, computational infrastructure, and research funding further marginalizes African voices within the global NLP community, hindering innovation and perpetuating dependency on external expertise. Decolonizing African NLP requires not only diversifying representation and inclusion within the field but also challenging the structural inequalities that perpetuate disparities in access and opportunity.

Furthermore, the neglect of African languages and cultural contexts in NLP research represents a missed opportunity for technological innovation and societal development. African languages encode unique ways of knowing and cultural perspectives that are essential for addressing local challenges and fostering sustainable development. By neglecting African languages in NLP research, we overlook valuable sources of knowledge and innovation, hindering efforts to address pressing issues such as healthcare, education, and governance. Decolonizing African NLP entails recognizing the intrinsic value of African languages and cultures and integrating them into the fabric of technological innovation (Adebara & Abdul-Mageed, 2022; Mboning Tchiaze & Assoumou, 2019).

In conclusion, the need for decolonization in African NLP is imperative. By challenging power dynamics, addressing bias and discrimination, and centering African languages and perspectives, we can pave the way for a more inclusive, equitable, and culturally responsive approach to NLP technology development in Africa. This requires collective action from researchers, practitioners, policymakers, and communities to dismantle colonial legacies and forge a path towards a more just and inclusive digital future for Africa and beyond.

4   ADDRESSING POWER DYNAMICS

Central to the decolonization of African NLP is the recognition and redressal of power dynamics inherent in technology development. Power imbalances manifest in various forms within the field, from unequal representation in research teams to disparities in access to resources and decision-making authority (Kgatle, 2022). Addressing these dynamics requires a concerted effort to redistribute power, amplify marginalized voices, and challenge Eurocentric frameworks that perpetuate colonial legacies (Craps, 2013).

One key aspect of addressing power dynamics in African NLP is diversifying representation within research teams and academic institutions. Historically, the majority of NLP research has been conducted in Western institutions, with limited participation from African scholars and technologists. By fostering greater diversity and inclusion within research teams, we can broaden perspectives, incorporate diverse experiences, and challenge dominant narratives that prioritize Western knowledge systems. This not only enriches the research process but also ensures that NLP technologies are more reflective of the needs and priorities of diverse communities.

Empowering local communities to participate in decision-making processes around technology development is essential for addressing power imbalances. Too often, decisions about NLP projects are made without meaningful input from the communities they purport to serve, leading to solutions that are disconnected from local contexts and needs. By engaging communities as partners rather than passive recipients of technology, we can ensure that NLP projects are more responsive, ethical, and sustainable. This requires building trust, fostering collaboration, and prioritizing the empowerment of marginalized voices within technology development processes.

In addition, challenging Eurocentric frameworks and methodologies is crucial for decolonizing African NLP. Many NLP models and algorithms are trained on datasets that reflect Western cultural norms and linguistic conventions, leading to biased and inaccurate results when applied to African languages and contexts. By developing alternative approaches that center African languages and perspectives, we can mitigate bias, improve accuracy, and promote greater inclusivity in NLP

technologies. This may involve adapting existing methodologies, co-designing solutions with local communities, and integrating indigenous knowledge systems into the research process.

In conclusion, addressing power dynamics is fundamental to the decolonization of African NLP. By diversifying representation, empowering local communities, and challenging Eurocentric frameworks, we can build a more equitable and inclusive approach to technology development that reflects the diverse needs and priorities of African societies. This requires a commitment to collective action, collaboration, and reflexivity within the NLP community to dismantle colonial legacies and foster a more just and equitable digital future for Africa and beyond.

5 COMBATTING DATA COLONIALISM

Data colonialism, a concept rooted in the exploitation and extraction of data from marginalized communities for the benefit of external actors, poses significant challenges to the decolonization of African NLP (Thatcher et al., 2016). In the context of Africa, data colonialism perpetuates historical patterns of exploitation and reinforces asymmetrical power relations, leaving African communities vulnerable to economic exploitation and cultural erasure. Combatting data colonialism requires a multifaceted approach that prioritizes data sovereignty, community ownership, and ethical data practices Calzati (2021).

At the heart of the struggle against data colonialism is the principle of data sovereignty, which asserts the rights of communities to control and govern their own data. In the African context, data sovereignty is essential for safeguarding against the extraction and exploitation of data by external actors, including multinational corporations and foreign governments. By asserting control over their data, African communities can determine how it is collected, used, and shared, thereby reclaiming agency over their digital narratives and asserting their right to self-determination.

Promoting community ownership of data is another crucial aspect of combatting data colonialism. Too often, data collected from African communities is commodified and monetized by external actors without regard for the interests or well-being of those from whom it was extracted. By empowering communities to own and manage their own data, we can ensure that the benefits of data-driven technologies accrue to those most affected by them, rather than to external actors seeking to profit from their exploitation. This may involve developing community-led data governance structures, implementing data-sharing agreements that prioritize community interests, and promoting participatory approaches to data collection and analysis.

Ethical data practices play a vital role in combatting data colonialism and promoting equitable data relationships. This includes ensuring informed consent and meaningful participation in data collection processes, as well as adopting transparent and accountable data management practices. By adhering to ethical principles such as fairness, transparency, and accountability, we can mitigate the risk of harm and exploitation associated with data colonialism, while promoting trust and reciprocity in data relationships (Mann & Daly, 2019).

In conclusion, combatting data colonialism is essential for the decolonization of African NLP. By prioritizing data sovereignty, community ownership, and ethical data practices, we can challenge the asymmetrical power relations that underpin data colonialism and foster more equitable and just data relationships. This requires a commitment to collective action, solidarity, and advocacy within the NLP community to dismantle colonial legacies and build a more equitable and inclusive digital future for Africa and beyond.

6 CASE STUDIES AND EXAMPLES

Examining real-world case studies and examples provides valuable insights into the practical application of decolonial approaches within the field of African NLP. These examples illustrate both the challenges and opportunities inherent in decolonizing NLP technologies, shedding light on the complexities of navigating power dynamics, data colonialism, and cultural contexts.

One compelling case study is the development of machine translation systems for African languages (Martinus & Abbott, 2019; Abbott & Martinus, 2018). Traditionally, machine translation systems have prioritized widely spoken languages such as English, French, and Spanish, while neglecting

many African languages. However, initiatives such as Masakhane—a community-driven project focused on advancing NLP research for African languages—are challenging this status quo by developing machine translation models specifically tailored to African languages (Gitau et al., 2023). By mobilizing local expertise, leveraging community resources, and adopting culturally sensitive approaches to translation, Masakhane exemplifies how decolonial approaches can promote linguistic diversity and inclusivity within NLP.

Another illuminating example is the use of NLP technologies to address social and environmental challenges in Africa. For instance, projects such as mPedigree[1], a platform that uses NLP to verify the authenticity of pharmaceutical products in Africa—are leveraging technology to combat counterfeit drugs and ensure access to safe and effective medication. By incorporating local knowledge, engaging with community stakeholders, and prioritizing ethical considerations, mPedigree demonstrates how NLP can be harnessed as a tool for social good while respecting local contexts and priorities.

Additionally, efforts to promote indigenous knowledge systems and cultural heritage through NLP technologies offer promising avenues for decolonization. Projects such as the African Storybook Initiative[2], which aims to create a digital repository of African stories in multiple languages—are preserving and promoting indigenous languages and cultural narratives through the use of NLP technologies. By providing access to diverse linguistic and cultural resources, the African Storybook Initiative empowers communities to reclaim agency over their digital identities and challenge dominant narratives that marginalize African voices.

These case studies and examples highlight the transformative potential of decolonial approaches within the field of African NLP. By centering African languages, perspectives, and priorities, and fostering collaboration and solidarity within the NLP community, we can forge a more equitable and inclusive future for technology development in Africa and beyond (Arakpogun et al., 2021). However, these examples also underscore the importance of continued research, advocacy, and community engagement to address the complex challenges of decolonizing NLP and ensure that technology serves the needs and aspirations of all people.

7 CONCLUSION

In conclusion, the imperative to decolonize African NLP is both urgent and compelling. Throughout this paper, we have explored the multifaceted dimensions of coloniality within the field of NLP, from power dynamics and data colonialism to cultural hegemony and linguistic marginalization. We have examined the pervasive influence of colonial legacies on technology development in Africa and the profound implications for equity, inclusion, and self-determination.

Decolonizing African NLP requires a concerted effort to challenge entrenched power dynamics, redistribute resources, and amplify marginalized voices within the field. It demands a reevaluation of dominant paradigms and methodologies, a commitment to centering African languages and perspectives, and a recognition of the intrinsic value of diverse knowledge systems and cultural narratives. It necessitates confronting data colonialism head-on, asserting data sovereignty, promoting community ownership, and fostering ethical data practices that prioritize the rights and well-being of African communities.

Yet, amidst these challenges lie opportunities for transformation and innovation. Initiatives such as Masakhane, mPedigree, and the African Storybook Initiative exemplify the potential of decolonial approaches to promote linguistic diversity, address social inequalities, and empower marginalized communities through NLP technologies. By building on these successes, fostering collaboration, and centering the voices and experiences of those most affected by technology, we can chart a course towards a more just, equitable, and inclusive digital future for Africa and beyond.

However, the journey towards decolonization is ongoing and requires sustained commitment from all stakeholders, including researchers, practitioners, policymakers, and communities. It demands humility, reflexivity, and a willingness to confront uncomfortable truths about the role of technology in perpetuating colonial legacies. It calls for solidarity, collaboration, and collective action to

---

[1] https://en.wikipedia.org/wiki/Mpedigree
[2] https://en.wikipedia.org/wiki/African_Storybook

dismantle oppressive systems and build new paradigms of technology development that prioritize justice, equity, and human dignity.

In the words of Kenyan writer and activist Ngũgĩ wa Thiong'o, "Decolonizing the mind is the most dangerous and difficult thing one can attempt." Yet, it is also the most necessary. As we embark on this journey of decolonizing African NLP, let us do so with courage, compassion, and conviction, knowing that the future we aspire to build is within our reach if we dare to imagine and create it together.

8   FUTURE DIRECTIONS

As we look to the future of decolonizing African NLP, several key directions emerge that warrant further exploration and action. These future directions encompass research, policy, and community engagement efforts aimed at advancing the decolonial agenda within the field of NLP and technology development more broadly.

Fostering Interdisciplinary Collaboration: Future efforts to decolonize African NLP should prioritize interdisciplinary collaboration, bringing together researchers, practitioners, and stakeholders from diverse fields such as linguistics, anthropology, sociology, and cultural studies. By integrating perspectives and methodologies from different disciplines, we can enrich our understanding of the complexities of coloniality and develop more holistic and inclusive approaches to technology development.

Building Capacity and Infrastructure: Investing in capacity building and infrastructure development is essential for advancing the decolonial agenda within African NLP. This includes supporting educational programs, research networks, and technical infrastructure that empower African scholars and technologists to participate meaningfully in NLP research and development. By strengthening local expertise and resources, we can reduce dependency on external actors and promote self-sufficiency and innovation within the African tech ecosystem.

Promoting Policy and Advocacy: Advocating for policy changes and institutional reforms is critical for creating an enabling environment for decolonizing African NLP. This includes advocating for data protection laws, open access policies, and funding mechanisms that prioritize equity, inclusion, and community ownership of data. By engaging with policymakers and advocating for policy changes, we can create the necessary regulatory frameworks and incentives to support decolonial approaches to technology development.

Centering Community Engagement: Centering community engagement and participatory approaches is essential for ensuring that decolonizing efforts are grounded in the needs, priorities, and aspirations of local communities. This involves fostering meaningful dialogue, building trust, and co-designing solutions with community stakeholders throughout the technology development process. By prioritizing community voices and experiences, we can ensure that NLP technologies are more responsive, equitable, and culturally relevant.

Continued Research and Reflection: Finally, continued research and reflection are crucial for advancing the decolonial agenda within African NLP. This includes critically examining the implications of colonial legacies on technology development, exploring alternative paradigms of knowledge production and dissemination, and documenting and sharing best practices and lessons learned from decolonial initiatives. By fostering a culture of reflexivity and continuous learning within the NLP community, we can adapt and evolve our approaches to better address the complex challenges of decolonization.

In conclusion, the future of decolonizing African NLP holds immense promise and potential. By embracing interdisciplinary collaboration, investing in capacity building and infrastructure development, advocating for policy changes, centering community engagement, and fostering continued research and reflection, we can forge a path towards a more just, equitable, and inclusive digital future for Africa and beyond. This future is within our grasp if we remain committed to the principles of justice, equity, and solidarity in our pursuit of decolonial technology development.

AUTHOR CONTRIBUTIONS

All authors contributed to the study design. NY played a major role in data collection, analysis, and interpretation. CX provided supervision. NY drafted the manuscript, with critical input from CX. All authors approved the final version for submission.

REFERENCES

Jade Z Abbott and Laura Martinus. Towards neural machine translation for african languages. *arXiv preprint arXiv:1811.05467*, 2018.

Rachel Adams. Can artificial intelligence be decolonized? *Interdisciplinary Science Reviews*, 46 (1-2):176–197, 2021. doi: 10.1080/03080188.2020.1840225. URL https://doi.org/10.1080/03080188.2020.1840225.

Ife Adebara and Muhammad Abdul-Mageed. Towards afrocentric NLP for African languages: Where we are and where we can go. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3814–3841, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-long.265. URL https://aclanthology.org/2022.acl-long.265.

Emmanuel Ogiemwonyi Arakpogun, Ziad Elsahn, Femi Olan, and Farid Elsahn. Artificial intelligence in africa: Challenges and opportunities. *The fourth industrial revolution: Implementation of artificial intelligence for growing business success*, pp. 375–388, 2021.

Nina Berman, Klaus Mühlhahn, and Patrice Nganang. *German colonialism revisited: African, Asian, and Oceanic experiences*. University of Michigan Press, 2014.

Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. Language (technology) is power: A critical survey of "bias" in NLP. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5454–5476, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.485. URL https://aclanthology.org/2020.acl-main.485.

Igor Calzada. Data co-operatives through data sovereignty. *Smart Cities*, 4(3):1158–1172, 2021.

Stefano Calzati. Decolonising "data colonialism" propositions for investigating the realpolitik of today's networked ecology. *Television & New Media*, 22(8):914–929, 2021.

Lucy Corkin. *Uncovering African agency: Angola's management of China's credit lines*. Routledge, 2016.

Nick Couldry and Ulises A Mejias. Data colonialism: Rethinking big data's relation to the contemporary subject. *Television & New Media*, 20(4):336–349, 2019.

Stef Craps. Beyond eurocentrism. *The future of trauma theory: Contemporary literary and cultural criticism*, 45, 2013.

Jamell Dacon. Towards a deep multi-layered dialectal language analysis: A case study of African-American English. In Su Lin Blodgett, Hal Daumé III, Michael Madaio, Ani Nenkova, Brendan O'Connor, Hanna Wallach, and Qian Yang (eds.), *Proceedings of the Second Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, pp. 55–63, Seattle, Washington, July 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.hcinlp-1.8. URL https://aclanthology.org/2022.hcinlp-1.8.

Toyin Falola. *Decolonizing African studies : knowledge production, agency, and voice*. Rochester studies in African history and the diaspora. University of Rochester Press, 2022. URL https://cir.nii.ac.jp/crid/1130856960299624341.

Christopher Farrands. Technology and the technical management of human resources: Prospects for sub-saharan african development into the new millennium. In *Africa at the Millennium: An Agenda for Mature Development*, pp. 60–78. Springer, 2000.

Anjalie Field, Su Lin Blodgett, Zeerak Waseem, and Yulia Tsvetkov. A survey of race, racism, and anti-racism in NLP. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 1905–1925, Online, August 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.acl-long.149. URL https://aclanthology.org/2021.acl-long.149.

José Antonio García-Díaz, Camilo Caparros-laiz, Ángela Almela, Gema Alcaráz-Mármol, María José Marín-Pérez, and Rafael Valencia-García. UMUTeam at SemEval-2023 task 12: Ensemble learning of LLMs applied to sentiment analysis for low-resource African languages. In Atul Kr. Ojha, A. Seza Doğruöz, Giovanni Da San Martino, Harish Tayyar Madabushi, Ritesh Kumar, and Elisa Sartori (eds.), *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pp. 285–292, Toronto, Canada, July 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.semeval-1.38. URL https://aclanthology.org/2023.semeval-1.38.

Peter Genger. Re-empowering african indigenous peacemaking approaches: Identifying the enabling possibilities from decolonization and indigenization discourses. *Indigenous knowledge systems and development in Africa*, pp. 131–157, 2020.

Catherine Gitau, Salomon Kabongo, Abiodun Modupe, Vukosi Marivate, Thapelo Sindane, Richard Lastrucci, and Isheanesu .J Dzingirai. Masakhane web: A machine translation platform for african languages, 2023. URL https://doi.org/10.21203/rs.3.rs-2646015/v1.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. The state and fate of linguistic diversity and inclusion in the NLP world. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault (eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 6282–6293, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.560. URL https://aclanthology.org/2020.acl-main.560.

Shose Kessi, Zoe Marks, and Elelwani Ramugondo. Decolonizing african studies, 2020.

Mookgo Solomon Kgatle. Power dynamics among pentecostal prophets in new prophetic churches: An african christocentric missional approach. *Pharos Journal of Theology*, 2022.

Philippe Le Billon. *The geopolitics of resource wars*. Routledge, 2017.

Monique Mann and Angela Daly. (big) data and the north-in-south: Australia's informational imperialism and digital colonialism. *Television & New Media*, 20(4):379–395, 2019.

Vukosi Marivate. Why african natural language processing now? a view from south africa# africanlp. 2020.

Laura Martinus and Jade Z Abbott. A focus on neural machine translation for african languages. *arXiv preprint arXiv:1906.05685*, 2019.

Elvis Mboning Tchiaze and Jules Assoumou. NTeALan - AI/NLP/NLU Platforms For Sharing and Leveraging African Language Resources For Education In Africa. International Conference Language Technologies for All (LT4All): Enabling Linguistic Diversity and Multilingualism Worldwide, December 2019. URL https://hal.science/hal-02520738. Poster.

Wilhelmina Nekoto, Vukosi Marivate, Tshinondiwa Matsila, Timi Fasubaa, Taiwo Fagbohungbe, Solomon Oluwole Akinola, Shamsuddeen Muhammad, Salomon Kabongo Kabenamualu, Salomey Osei, Freshia Sackey, Rubungo Andre Niyongabo, Ricky Macharm, Perez Ogayo, Orevaoghene Ahia, Musie Meressa Berhe, Mofetoluwa Adeyemi, Masabata Mokgesi-Selinga, Lawrence Okegbemi, Laura Martinus, Kolawole Tajudeen, Kevin Degila, Kelechi Ogueji, Kathleen Siminyu, Julia Kreutzer, Jason Webster, Jamiil Toure Ali, Jade Abbott, Iroro Orife, Ignatius Ezeani, Idris Abdulkadir Dangana, Herman Kamper, Hady Elsahar, Goodness Duru, Ghollah Kioko, Murhabazi Espoir, Elan van Biljon, Daniel Whitenack, Christopher Onyefuluchi, Chris Chinenye Emezue, Bonaventure F. P. Dossou, Blessing Sibanda, Blessing Bassey, Ayodele Olabiyi, Arshath Ramkilowan, Alp Öktem, Adewale Akinfaderin, and Abdallah Bashir. Participatory research for low-resourced machine translation: A case study in African languages. In

Trevor Cohn, Yulan He, and Yang Liu (eds.), *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 2144–2160, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.findings-emnlp.195. URL https://aclanthology.org/2020.findings-emnlp.195.

Xuan-Phi Nguyen, Sharifah Mahani Aljunied, Shafiq Joty, and Lidong Bing. Democratizing llms for low-resource languages by leveraging their english dominant abilities with linguistically-diverse prompts, 2023.

Artwell Nhemachena, Nokuthula Hlabangane, and Maria B Kaundjua. Relationality or hospitality in twenty-first century research? big data, internet of things, and the resilience of coloniality on africa. *Modern Africa: Politics, History and Society*, 8(1):105–139, 2020.

Iroro Orife, Julia Kreutzer, Blessing Sibanda, Daniel Whitenack, Kathleen Siminyu, Laura Martinus, Jamiil Toure Ali, Jade Abbott, Vukosi Marivate, Salomon Kabongo, Musie Meressa, Espoir Murhabazi, Orevaoghene Ahia, Elan van Biljon, Arshath Ramkilowan, Adewale Akinfaderin, Alp Öktem, Wole Akin, Ghollah Kioko, Kevin Degila, Herman Kamper, Bonaventure Dossou, Chris Emezue, Kelechi Ogueji, and Abdallah Bashir. Masakhane – machine translation for africa, 2020.

Nicole Redvers, Michael Yellow Bird, Diana Quinn, Tyson Yunkaporta, and Kerry Arabena. Molecular decolonization: An indigenous microcosm perspective of planetary health. *International Journal of Environmental Research and Public Health*, 17(12), 2020. ISSN 1660-4601. doi: 10.3390/ijerph17124586. URL https://www.mdpi.com/1660-4601/17/12/4586.

Lane Schwartz. Primum Non Nocere: Before working with Indigenous data, the ACL must confront ongoing colonialism. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 724–731, Dublin, Ireland, May 2022. Association for Computational Linguistics. doi: 10.18653/v1/2022.acl-short.82. URL https://aclanthology.org/2022.acl-short.82.

Edward Shizha. *African Indigenous Perspectives on Technology*, pp. 47–62. SensePublishers, Rotterdam, 2016. ISBN 978-94-6300-515-9. doi: 10.1007/978-94-6300-515-9_4. URL https://doi.org/10.1007/978-94-6300-515-9_4.

Raphael Tang, Xinyu Zhang, Jimmy Lin, and Ferhan Ture. What do llamas really think? revealing preference biases in language model representations, 2023.

Jim Thatcher, David O'Sullivan, and Dillon Mahmoudi. Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, 34(6):990–1006, 2016.

Arafaat Amin Valiani. Frontiers of bio-decolonization: Indigenous data sovereignty as a possible model for community-based participatory genomic health research for racialized peoples in postgenomic canada. *Genealogy*, 6(3), 2022. ISSN 2313-5778. URL https://www.mdpi.com/2313-5778/6/3/68.

Cheng Xu and Nan Yan. AROT-COV23: A dataset of 500k original arabic tweets on COVID-19. In *4th Workshop on African Natural Language Processing*, 2023. URL https://openreview.net/forum?id=aUZhVQBl2W.