Embedded Coloniality in Language Data, Algorithms, and Software

Anonymous ACL submission

Abstract

Coloniality, the continuation of colonial harms beyond "official" colonization, has pervasive effects across society and scientific fields. Natural Language Processing (NLP) is no exception to this broad phenomenon. In this work, we argue that coloniality is implicitly embedded in and amplified by NLP data, algorithms, and software. We formalize this analysis using Actor-Network Theory (ANT): an approach to understanding social phenomena through the network of relationships between human stakeholders and technology. We use our Actor-Network to guide a quantitative survey of the geography of different phases of NLP research, providing evidence that inequality along colonial boundaries increases as NLP builds on itself. Based on this, we argue that combating coloniality in NLP requires not only changing current values but also active work to remove the accumulation of colonial ideals in our foundational data and algorithms.

1 Introduction

While European colonization, the sovereignty of European nations over non-European nations, has been dismantled on paper in most of the world[1], it still mars modern society. **Coloniality**, the continuation of harms against the colonized and centering of the ideals of colonizing powers, is pervasive across geographic regions, intellectual domains, and sectors (Said, 1978; Smith, 1999; Quijano, 2000; Maldonado-Torres, 2007; Boshoff, 2009; Chakrabarti, 2010; Wengraf, 2018; Alvarado Garcia et al., 2021; Ferdinand, 2021), including AI broadly (Mohamed et al., 2020; Birhane, 2020).

Language studies specifically have a stark colonial history. Up until the 20th century, the linguistics institutions of Non-European languages were led by European scholars, often supported directly by colonial regimes (Errington, 2001) which suppressed the use of those same languages by their native speakers (Wa Thiong'o, 1992). Prior works have argued that NLP continues to perpetuate methodological coloniality, especially in research on African and Indigenous Languages (Bird, 2020; Schwartz, 2022; Ògúnrẹ̀mí and Samuel, 2023; Mager et al., 2023).

These works also show coloniality is not simply a matter of linguistic diversity. As illustrated in Figure 1, both languages studied *and* researcher affiliations reflect Western Eurocentrism. The study of new languages can often reflect the shifting interests of existing powers, called interest convergence (Bell Jr, 1980; Ogbonnaya-Ogburu et al., 2020). For example, in times of geopolitical conflict with the US, languages of what was then the Soviet Union (1960-1980) (Léon, 2021) and later the Middle East (2000-2010) (Farghaly and Shaalan, 2009) experienced a surge in research, without a proportional increase in the representation of researchers affiliated with those regions.

As many believe that NLP technologies are on the precipice of enacting major societal (Bommasani et al., 2021) and economic change (Eloundou et al., 2023), understanding inequity in people's agency to shape NLP is of increasingly broad importance. Otherwise, systems such as large language models (LLMs) are likely to expand harms such as algorithmic discrimination, exclusion, and stereotyping (Bender et al., 2021; Weidinger et al., 2022; Hovy and Spruit, 2016; Cheng et al., 2023).

In this paper, we focus on how coloniality can be connected to unstructured data, annotations, models, and software in NLP. In turn, these technologies reinforce and amplify coloniality beyond the social systems that created them. This view of technology and objects as agents that resist or create change in sociotechnical systems is called "**materiality**" (Winner, 1980; Appadurai, 1988; Keane, 2003; Hodder, 2012).

To formally guide our analysis, we leverage

[1] There are still 17 territories colonized by the United States, the United Kingdom, and France.

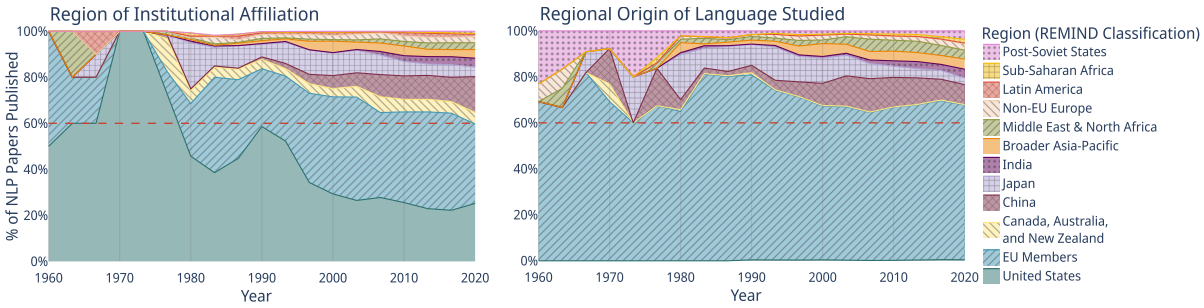


Figure 1: Natural Language Processing (NLP) research grouped by the regional affiliation of authors and regional origin of languages studied using the REMIND region classification (Baumstark et al., 2021). Across 7 decades of NLP research, institutions based in Western Europe and British settler-colonial states have always published over 60% of NLP papers. While science broadly follows this dynamic (Nature, 2014), in NLP this shapes the research itself to be Eurocentric, across the same period over 60% of research focused on Western European languages.

Actor-Network theory (Latour, 2005) (ANT), a conceptual framework that allows us to relate people and technology jointly. ANT is widely used in Science and Technology Studies, including for digital and data-driven technologies (Bowker and Star, 1999; Stanforth, 2006; Kumar and Rangaswamy, 2013; Frauenberger, 2019; Selbst et al., 2019). In ANT, both human and technological actors are connected via their relationships in a so-called Actor-Network. This can then be used to identify people, organizations, and technologies that stabilize and destabilize power dynamics in the network.

To support understanding of coloniality in NLP, we contribute the following:

1. **Survey of the NLP Actor-Network:** We survey prior literature that documents technological artifacts and their production process. We then synthesize these works into an Actor-Network for NLP, shown in Figure 2, covering data production, annotation, modeling, and deployment.
2. **Interpreting the Material Expression of Coloniality in NLP:** We analyze the direct effects of coloniality on technological actors with this Actor-Network. We ground this analysis in a quantitative study of all *CL publications up to September 2022, described in Section 3, to measure regional involvement in different stages of NLP development.
3. **Qualitative Studies of Coloniality in NLP:** We use qualitative case studies to highlight that our quantitative analysis is an underestimate of impacts since geographic borders and languages are only surface-level measures.

We argue that reliance on high-end hardware and increasingly large datasets stabilize inequality in regional agency. Individuals in the US, Canada, Australia, New Zealand, and EU Member States make up over 50% of the research community involved in annotation, modeling, and deployment despite making up less than 20% of those who produce language data online.

Liboiron (2021) notes that coloniality "*can be maintained by good intentions and even good deeds.*" Through our focus, we argue that these material aspects of coloniality have pervasive downstream effects on NLP researchers, even if we assume good intentions. Therefore, we must actively work to reshape these technological foundations of NLP according to communal values.

2 Building The Actor-Network

We develop our Actor-Network, displayed in Figure 2, by surveying works that document NLP data, modeling, and deployment processes.

2.1 Unlabeled Digital Data Production

The foundation of NLP is data upon which datasets are constructed, systems are trained and tested, and products operate. Often, this is sourced from the Internet and it must be digitized in some fashion.

Infrastructure provided by Internet Service Providers is central for the data in most NLP datasets. This barrier, which is largely dependent on geography (Pimienta et al., 2009), limits the representativeness of the raw data distribution. Even for offline textual data, the language needs support from both hardware, such as keyboards (Mullaney, 2017), and software, such as encoding schemes (Dombrowski, 2020), to allow for the production and consumption of digital text (Paolillo et al., 2005) or be scanned and extracted from non-digital resources (Rijhwani et al., 2020).

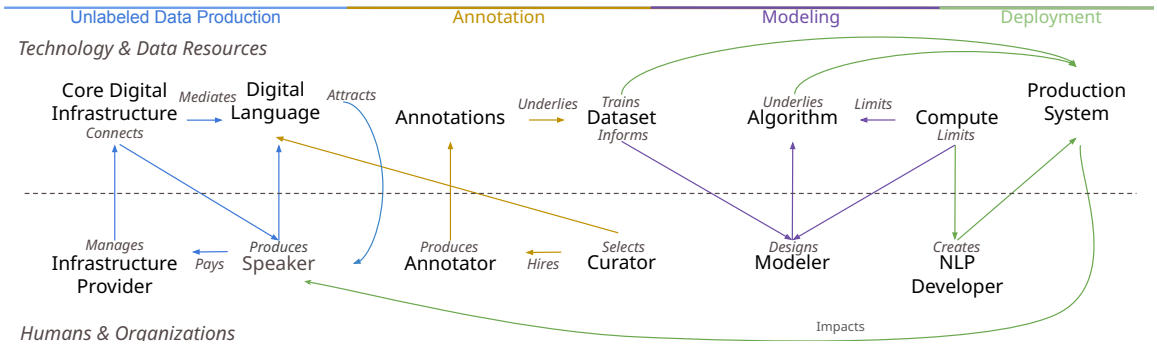


Figure 2: The Actor-Network of NLP Development and Deployment. Unlabeled data production process documented in Pimienta et al. (2009) and Kornai (2013). Annotation process documented in Bender and Friedman (2018). Modeling process documented in Ethayarajh and Jurafsky (2020) and Hooker (2021). Deployment and impacts documented in Hovy and Spruit (2016) and Mitchell et al. (2019).

Beyond access, social factors influence discrepancies in the production of language data (Hovy and Yang, 2021; Hovy and Spruit, 2016). The strong network effects of content production make the size of existing online communities a stronger predictor of language vitality online than raw population numbers (Kornai, 2013). As speakers are attracted to the web by existing websites, they themselves are likely to become content producers. Translation does not resolve this issue, as the quality of software (Kornai, 2013) and the investment in digital literacy (Hargittai, 2001) create their own barriers to adoption and content production.

2.2 Annotated Datasets as Stores of Value

In addition to unlabeled data, NLP systems often rely on curated examples of language with additional annotations. From the outset, the process of curation is affected by the raw data distribution. As the process of data curation and annotation is repeated, curators often use the same distribution of data (Raji et al., 2021) leaving others systematically excluded (Hovy and Spruit, 2016).

In developing standards of documentation for NLP datasets, Bender and Friedman (2018) carefully connect the social aspects of language dataset creation to the dataset itself. This is an inherently materialist perspective — this documentation is necessary because the resource propagates these social forces. Speakers who produce raw data, curators who select raw data, and annotators who produce individual annotations, all impact the resulting datasets.

Curators are especially notable in this process as, beyond curation, they often recruit annotators and design annotation guidelines. This dynamic has increased as crowdwork has become prevalent in the field (Shmueli et al., 2021), especially for large-scale Reinforcement Learning From Human-Feedback (Ouyang et al., 2022; Touvron et al., 2023). This leaves curators as the primary authority on the guidelines, labels, and formats that underly annotated datasets. By deciding how the near infinite space of language is compressed into fixed categories, the ideals implicit in this structure are applied not only to the data that they have collected but to all data they feel it represents (Todd, 2016).

2.3 Social Incentives in Modeling Work

One major difference between the closely related fields of NLP and Computational Linguistics is the focus on empirical results. A famous, albeit of contested exact phrasing, quote from an NLP pioneer succinctly captures this value difference: "*Every time I fire a linguist, the performance of the system goes up.*" (Hirschberg, 1998). This shift in priorities was shaped by required participation in shared tasks during the MUC (Chinchor, 1991; Sundheim, 1995) and ATIS (Hemphill et al., 1990) workshops in the late 1980's and early 1990's. These were central events in the field at the time since grantees funded by the US Defense Advanced Research Projects Agency (DARPA) were encouraged to attend (Anderson et al., 2012). As participants in these events established and led future areas of NLP, benchmarks and shared tasks became a major driver of NLP model development (Strassel and Cole, 2006; Paroubek et al., 2007; Wang et al., 2018; Raji et al., 2021; bench authors, 2023).

However, defining and ranking by "performance" requires reducing the qualitative features of a model to a minimal set of sortable metrics. Inevitably, some aspects of true performance are lost. Ethayarajh and Jurafsky (2020) use game

theory to argue that this rationally dictates the priorities of model developers. Optimizing widely reported dataset metrics can lead to publicity and career rewards. These incentives can override other dimensions of practical utility to speakers (Ruder et al., 2022). Practically, modeling researchers report data availability as a major factor in deciding research directions (Thakkar et al., 2022) due to the relatively low prestige and high difficulty of data work (Sambasivan et al., 2021).

Furthermore, Hooker (2021) highlights that algorithm design has a second non-theoretical limiting factor — hardware. Algorithms that are mathematically or linguistically sound often fail if they are not well-supported by existing hardware. Even the now dominant neural networks exploded in popularity primarily after high-quality frameworks for GPU support were released (Bergstra et al., 2010). More recently, *Large* Language Models have developed alongside hardware to support them (Saphra et al., 2023), leading to increasing centralization (Bommasani et al., 2023).

Benchmarks and hardware, both non-human actors, thereby influence the design process of model developers. These constraints increase as NLP becomes more machine learning driven (Manning and Schutze, 1999; Sevilla et al., 2022) and especially reliant on larger datasets (Gao et al., 2020; Birhane et al., 2022b; Wang et al., 2018) and models (Strubell et al., 2019; Brown et al., 2020; Henderson et al., 2020; Patterson et al., 2021).

2.4 NLP Deployment Realizes Impact

Thus far we have focused on social forces shaping NLP, but NLP research leads to software products and tools that impact broader society. Similar to data statements, Mitchell et al. (2019) provides guidelines for documenting how Machine Learning model development should dictate deployment. Released models, their functionality, and their allowed use shape possible software products.

However, when multiple models are integrated into products, they transform these technical functionalities into non-technical impacts on those who use the product or to whom it is applied. As noted in Hovy and Spruit (2016), these systems impact the broader community that uses a language, regardless of whether they were included in the annotation or algorithm development process.

These applications are not asocial. They can have positive impacts such as creating economic opportunities (Brynjolfsson and Mitchell, 2017; Brynjolfsson et al., 2019), informing public policy (Jurgens et al., 2017a), and aiding education (Loukina et al., 2019; Wang and Demszky, 2023) or negative impacts through dual use such as harming privacy (Jurgens et al., 2017b), distributing harmful knowledge (Shaikh et al., 2023), and generating misinformation (Zhou et al., 2023). Through deployment, inequities in the social systems that shape NLP are replicated through technological harms and allocative inequities.

3 Quantitative Actor-Network Analysis

While theoretical analysis of the Actor-Network itself can create interesting hypotheses, one benefit of looking at material impacts is that measured. At the scale of the entire field, we believe there are two key universal traits in NLP works: the social environment that shapes the research and the language the research studies. To avoid generalizing about authors by identity, we approximate "social environment" via the institutional affiliations and, by proxy, regional affiliation of each work.

We analyze these traits across the full set of *CL papers released on the ACL Anthology up until September 2022 from the ACL-OCL-Corpus (Rohatgi et al., 2023). See Appendices A and B respectively for details on how we extract author country affiliations and language of interest for authors.

In order to assess how geographic diversity changes at different stages of the NLP research and deployment, we use heuristics, described in detail in Appendix C, to representative samples of each phase associated with our Actor-Network. For this analysis, we restrict these to papers published since the responsible NLP checklist was released to avoid temporal shifts between groups.

3.1 Annotation & the Research Agenda

As discussed in our development of the Actor-Network, the process of repeatedly resampling data magnifies its biases. In the "Annotation" section of Figure 3, we see that over 60.3% of papers from LREC come from Europe alone. This is likely influenced by LREC being organized by the European Language Resource Association LREC and primarily hosted within Europe. The annotated resources produced for LREC influence subsequent research agendas and further solidify the initial biases into the foundations of the field.

For almost any interest area, English datasets are more likely to already exist to test many different

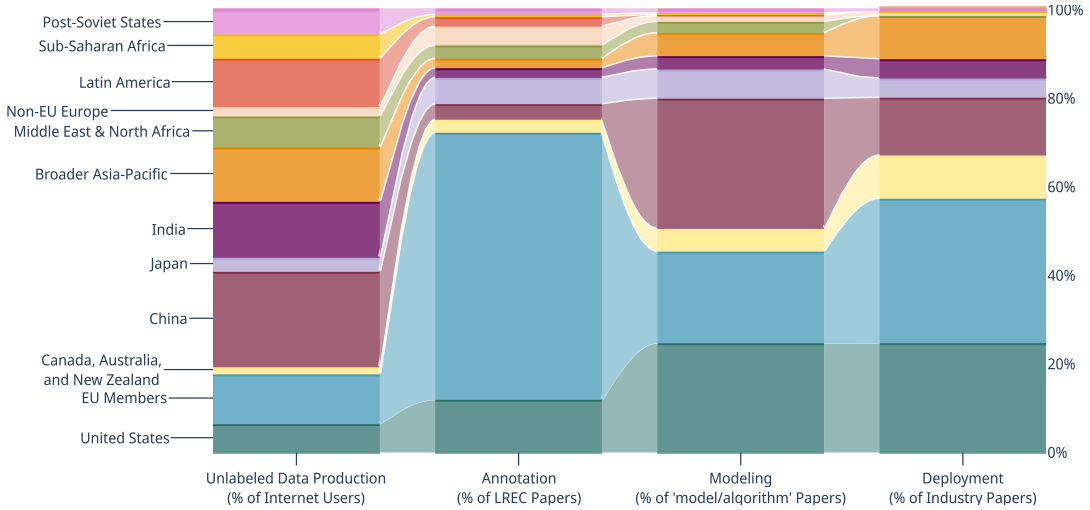


Figure 3: Geographic distribution of the people involved with each phase of the NLP pipeline, from raw data to deployed model, in papers since 2015. Representation of heavily colonized regions, such as Latin America and Sub-Saharan Africa, dwindles throughout the NLP pipeline.

modeling techniques or tasks (Faisal et al., 2022; Longpre et al., 2023). If a researcher's interests are truly "language-agnostic" or even if they have a slight preference to work on other languages, existing English resources may bias them to work on English to focus more directly on modeling (Thakkar et al., 2022) and avoid annotation where the path to tangible results requires greater time and financial investment (Sambasivan et al., 2021).

This process centers modeling in NLP on languages of the Global North where datasets already exist. Historically, NLP has termed systems "language-agnostic" even when the performance is only tested on English (Bender, 2011). When systems are not tested on typologically and geographically diverse languages but are termed "language-agnostic" or simply "multilingual", it perpetuates the idea that all languages are substitutable. This is reductive, recreating colonial scientific tendencies to "*[anatomize] and [melt] human entities as if they were so much inert matter*" (Said, 1978).

Positively, as shown in Joshi et al. (2020) and reproduced for our corpus in Figure 4a, linguistic diversity[2] in NLP is increasing. However, diversity metrics are negatively biased for small populations (Konopiński, 2020) and the increase coincides with increased NLP publishing volume. By comparison, Shannon Equitability, which normalizes the entropy used in Joshi et al. (2020) by the maximum possible entropy (Sheldon, 1969), has not significantly increased since the publication of the first ACL Proceedings. This inequality is clearly stated through the Gini Coefficient, a standard economic inequality metric, shown on the right in Figure 4b: Inequality between the study of languages has never left the range of "extreme inequality[3].

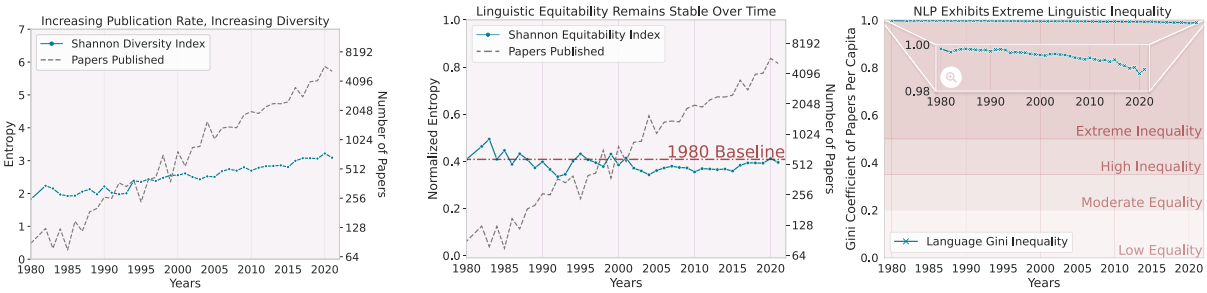
This challenges diversity alone as an indicator of equity, especially in regard to credit allocation. Data production for left-behind languages generally has a smaller audience of researchers than similar work done in English or other high-resource languages. This can be seen empirically through the relatively small clusters of publishing authors for most languages (Joshi et al., 2020). Similarly, North American and European research receives systematically higher citations (Rungta et al., 2022). In the "Modeling" section of Figure 3, we see that while representation of China expands (29.4% of Modeling papers, +25.9 from Annotation), South American (0.33% of Modeling papers, -1.79 from Annotation) and African (0.28% of Modeling papers, -0.26 from Annotation) representation dwindles to near-zero despite these regions making up 16.4% of all internet users.

3.2 Hardware & the Control of Deployment

Central to coloniality in NLP is the idea that stakeholders are often subject to NLP systems without the agency to change or reject them (Bird, 2020; Ògúnrèmí and Samuel, 2023; Blanco et al., 2023). In the "Deployment" section of Figure 3, we ap-

[2]We follow Joshi et al. (2020), quantifying diversity using the Shannon Entropy of language mentions. Broadly, this usage is called the Shannon Diversity Index(Magurran, 2021).

[3]Based on Gini Inequality categories used for income inequality by the UN ILO (Luebker, 2010)



(a) Linguistic diversity on the scale of the maximum and minimum possible entropy.

(b) Linguistic equality, measured by both Shannon Equitability (left) and Gini Inequality (right), on the scale of maximum and minimum measure values.

Figure 4: Diversity and equitability of languages mentioned in *CL publications for each year since the first Proceedings of the ACL. Metrics based on Shannon Entropy are compared to the number of papers published each year in NLP. See Appendix B for details on how we counted mentions in papers and computed each metric.

proximate this using papers published in the Industry track and workshops focused on NLP "in practice" . There are no papers from researchers or companies based in the Middle East, North Africa, Sub-Saharan Africa, or South America in these deployment-focused venues.

However, NLP is not the first colonial technology deployed across the globe and nations that suffered colonization have a history of adapting technologies used in service of colonialism into locally valuable tools, a process called appropriation (Fanon, 1959; Srinivasan, 2006; Escobar, 2011). This requires that people can own, modify, and reject technology according to their own value systems and desired usage. Recently, NLP has seen many appropriative initiatives that develop with and for historically colonized communities such as AmericasNLP (Mager et al., 2018, 2021; Ebrahimi et al., 2022, 2023), Masakhane (Orife et al., 2020; Nekoto et al., 2020; Adelani et al., 2021, 2022a,b), and AI4Bharat (Kakwani et al., 2020; Mhaske et al., 2023). Nekoto et al. (2020) is a gold standard in how this process can empower a community. Their work is led by speakers of target low-resourced African languages and centers on providing agency to all actors rather than just the technologists involved.

Appropriative efforts such as this can face a major material barrier — the low-resource double bind (Ahia et al., 2021). Even when data is accessible, state-of-the-art models often assume access to high-performance computing systems, such as the increasing reliance on Graphics Processing Units (GPUs). Access to even low-end GPUs is not universal. Over 50% of GPU usage in recent *CL publications is from just two countries, China and the United States. While both nations are large producers of NLP research in general, a paper that uses a GPU is 22.7% more likely to be from these two nations than a randomly sampled paper in the same time period. We show the full distribution of GPU use in Appendix C.1.

Even if access is provided through APIs, where there are no fixed costs, the appropriative process is subjected to the limitations laid out by providers. Services owned and developed in North America and Europe are often poorly designed for the historically colonized while undermining local innovation (Birhane, 2020). With recent NLP models, we have already seen more costly pricing and weaker safety systems for low-resource language speaking communities (Ahia et al., 2023; Yong et al., 2023).

4 Case Studies In Coloniality

We present two case studies to illustrate that quantitative measures capture only a subset of the impacts of coloniality on technological artifacts.

4.1 Disparities for Minority English Dialects

Key Material Actors: Digital Language, Datasets

While the Black population of the United States is often not regarded as colonial subjects, Black people in the U.S. are subject to internal colonialism[4]. During chattel slavery, Black people in the US were forced to adopt English and stop speaking indigenous African languages, mixing grammar patterns, pronunciations, and expression of various West African languages with 17th century British

[4] Gutiérrez (2004) defines internal colonialism as a system of exploitation within the oppressor state characterized by four key relationships of domination: the forced entry of the oppressed group to the dominant society; the transformation, or destruction of the value systems of the colonized; the administration of the colonized population by the oppressor nation; and racism towards the colonized population.

English (Winford, 2015). This resulted in African American Vernacular English (AAVE) also called African American English (AAE) or African American Language (AAL).

Enslaved Africans were barred from writing, creating a centuries-long oral tradition (Winford, 2015). In recent history, AAVE has been excluded from institutions such as media, literature, and education (Jackson, 1997). This has created a lack of unlabeled textual resources for AAVE, except for recently collected social media data from Black users (Blodgett and O'Connor, 2017). Even these are collected without consent or speaker interaction. As a result, AAVE is decontextualized creating risks of mockery and cultural appropriation from representations of AAVE available online (Ronkin and Karn, 1999; Smokoski, 2016). However, they are often the only naturally occurring data to train and evaluate NLP models for AAVE (Blodgett and O'Connor, 2017; Jørgensen et al., 2016).

The systematic exclusion of AAVE from both unlabeled and annotated data has resulted in models that may not accurately represent or apply to the Black population, but are applied without contextualization. Multiple prior works have shown that AAVE is more likely to be falsely classified as hate speech or toxic speech (Harris et al., 2022; Sap et al., 2019; Davidson et al., 2019; Halevy et al., 2021) and as having negative sentiment in sentiment analysis (Groenwold et al., 2020a). One downstream influence of this is the negative impact on the social media experiences of Black users, who are more likely to have their social media posts removed (Haimson et al., 2021; Harris et al., 2023).

Limited access to representative unlabeled data and qualified human annotators makes improving or even identifying these issues challenging. While interventions have been attempted to mitigate biases towards AAVE in annotation (Sap et al., 2019; Ziems et al., 2022; Dacon, 2022) none are a cohesive solution, especially due to the regional and generational variation of AAVE (Hinton and Pollock, 2000; King, 2020). Without fundamentally reshaping annotation procedures with feedback and data offered by AAVE speakers (Mengesha et al., 2021), new systems will perpetuate the issue even if researchers follow best practice mitigations.

The visibility of African American English within NLP (Jørgensen et al., 2016; Blodgett and O'Connor, 2017; Dorn, 2019; Groenwold et al., 2020b; Ziems et al., 2022) compared to other similarly situated languages such as Jamaican Patois (Armstrong et al., 2022), Haitian Kreyol (Ponti et al., 2020; Lent et al., 2021, 2022), Australian Aboriginal English (Zwarts and Dras, 2007; Ziems et al., 2023) is, in itself, a reflection of power of the US in NLP research. Furthermore, the Black communities in the United States that are most impacted are often not involved in AAVE NLP research. While some colonized groups have their languages excluded, others such as AAVE speakers, have language technologies imposed on them.

4.2 English Centric Multilingualism

Key Material Actors: Datasets, Algorithms, Hardware

Multilingual Language Models (Pires et al., 2019; Conneau et al., 2019; Xue et al., 2021) have successfully driven an increased research interest in NLP beyond English (Joshi et al., 2020). The growing popularity of such models has itself motivated a wave of cross-lingual benchmarks which annotate a single task over several languages (Conneau et al., 2018; Yang et al., 2019; Nivre et al., 2018; Pan et al., 2017; Artetxe et al., 2020b; Lewis et al., 2020; Clark et al., 2020; Zweigenbaum et al., 2018; Artetxe and Schwenk, 2019; Li et al., 2021; FitzGerald et al., 2022).

When researchers look to train multilingual systems, they usually adopt algorithms developed for the data-rich English context. These systems have data requirements far beyond what any single human consumes (Warstadt et al., 2023). Furthermore, reaching new performance heights seems to require exponentially more data with each new model (Kaplan et al., 2020; Hoffmann et al., 2022). In pursuit of data to meet this scale, multilingual models often sample as much unlabeled data as possible from the internet (Kreutzer et al., 2022). The impact of coloniality on unlabeled data production means that Western European languages, especially English, are over-represented compared to their relative population size (Chung et al., 2023). By this process, even multilingual models are demonstrably skewed towards Western, and especially American, values and phrasing (Johnson et al., 2022; Arora et al., 2023; Naous et al., 2023; Papadimitriou et al., 2023; Durmus et al., 2023).

Ideally, benchmarks would capture cultural discrepancies. However, data curators often translate existing annotated data (Artetxe et al., 2020a), likely aiming to extend an English-centric resource with accepted value to new languages. These datasets cover many languages, but the names, places, and other cultural concepts are transferred

into new languages. For example, out of the 47 Wikipedia articles in the XQuaD (Artetxe et al., 2020a) test set 50% are about Europe and North America, 36% are location agnostic, and only 15% are about the rest of the world.

This is common practice: 8 out of 12 tasks across in the widely used multi-task XTREME (Hu et al., 2020) and XTREME-R (Ruder et al., 2021) benchmarks use data translated from English. However, such datasets may struggle to capture cultural misalignment and may, in fact, reward systems that replicate, especially in the range of tasks where cultural understanding has been shown to influence accuracy (Mohammad et al., 2016; Smith et al., 2016; Asai et al., 2021; Bauer et al., 2023; Lee et al., 2023; Akinade et al., 2023).

5 Looking Forward

We have shown that the power dynamics of NLP become increasingly stark across colonial boundaries as the broader social inequities of coloniality are exacerbated by NLP specific technological barriers. These further prevent the historically colonized from developing empowering interventions. This inequality of choice (Sen, 1995, 2001) leaves the historically colonized subject to the choices of technologists in the Global North. In our analysis, two of these barriers stand out as technical challenges where researchers can expand agency in the NLP actor network listed below.

Reducing the low-resource double bind (Ahia et al., 2021) is a key short term focus. Hardware limitations create major barriers for those in the Global South to appropriate language technology for their own use (Aji et al., 2022). Appropriation of technology (Dix, 2007) is a partial solution that is well-established in cultures that survived colonialism, such as Jugaad in India (Rangaswamy and Densmore, 2013) or Gambiarra in Brazil (Boufleur, 2006). Researchers who view themselves as "purely technical" can enable appropriation of existing technology according to local value systems without forcing technological solutionism by developing more efficient methods. The expansion of agency and choice can itself be seen as technological development (Sen, 2001; Kleine, 2009), even if the outcome is the rejection of language technologies by affected stakeholders. Notably, appropriation alone is insufficient as it is an act of survival under oppressive power structures rather than an indicator of equality (Bar et al., 2016).

Methods that reduce the reliance on unfathomably large datasets to produce high-quality language technology, such as Adelani et al. (2022a) and Ògúnrèmí et al. (2023), help break the inheritance of their problematic foundations (Birhane and Prabhu, 2021). Data efficient learning allows curators to carefully design datasets in accordance with community-aligned values, rather than simply minimizing the biases of online data. Creating room for value-centered curation is key to re-imagining datasets as a rudder which helps direct NLP as a field (Bommasani, 2023) rather than as a measure of opaque notions of progress (Raji et al., 2021).

Technical work is not enough; we echo the calls of prior works (Ògúnrèmí and Samuel, 2023; Bird, 2020; Schwartz, 2022; Bird, 2022) that the NLP research community to make room for methodologies outside colonial traditions (Smith, 1999) and engage in participatory research (Nekoto et al., 2020). This may sometimes require more time for discussion to resolve conflict between viewpoints. However, this process is essential both ethically and to build systems that are effective in nuanced global contexts (Nkemelu et al., 2022).

Achieving linguistic diversity in NLP research is not, in itself, inclusion. Stakeholders from the Global South are often involved as annotators for their languages but excluded from the financial or reputational gains that authors receive from high-profile NLP projects (Shmueli et al., 2021)[5], and the Wall Street Journal. This perpetuates colonial methodologies in research: "research projects are designed and carried out with little recognition accorded to the people who participated" (Smith, 1999). Participatory research is not a panacea for the malcontents of NLP if it does not provide agency and ownership of the language technology developed to the stakeholders subject to resulting systems (Birhane et al., 2022a). Beyond technosolutionism, we must consider and discuss with stakeholders when NLP may exacerbate harm.

Our code of ethics binds us to, regardless of intent, "ensure that all harm is minimized" (Gotterbarn et al., 2018)[6]. By understanding the ties between NLP and colonial ideals, we see that doing so is an active process of undoing the ways in which harm is inextricably tied to the foundations of our field.

[5] Reports on exploitation of annotators in the Global South for NLP in the MIT Tech Review, TIME

[6] The ACL officially follows the ACM Code of Ethics.

6 Limitations

In developing our actor network, we focus on the relationships which exist pervasively across NLP. As the NLP research community is large with many focus areas, this generalized framework is likely to omit important actors which are relevant within specific contexts of NLP. When applying ANT to specific NLP tasks, we recommend using our framework for common elements, but moving towards defining concrete context specific key actors.

Furthermore, due to the expansiveness of European colonialism, coloniality has affected many diverse groups. Our work looks to prior work across these perspectives, but readers should not take our work as an indication that the desires of these communities are homogeneous. As we explore in our case study of African American Vernacular English, for some communities, language technologies are neither desired nor beneficial. Inclusion in the actor network should also be reflected as the right to reject language technology entirely.

Finally, while our work focuses on the material expressions of coloniality, this is in addition to, not suggesting the lack of, ideological coloniality highlighted in prior work. This work was inspired by prior works by Ògúnrẹ̀mí and Samuel (2023), Mohamed et al. (2020), Schwartz (2022), Bird (2020) Mager et al. (2023), and Blanco et al. (2023). This focus should not be taken to indicate that ideological coloniality has been solved in NLP. Ideological coloniality continues to be pervasive and serves to increase material inequities.

7 Ethical Considerations

This work uses the term "the historically colonized" to draw attention to the fact colonialism continues to affect those who are not presently under an official colonial regime. The use of this term may give the false impression that we live in a postcolonial context. The peoples of Anguilla, Bermuda, the Virgin Islands, the Cayman Islands, Malvinas, Montserrat, Saint Helena, Turks and Caicos, Gibraltar, American Samoa, French Polynesia, Guam, New Caledonia, Pitcairn, and Tokelau are still colonized[7]. In addition to these officially colonized nations, many historically colonized nations maintain neocolonial relationships, where former occupying nations still exert influence through economic and political power. While close analysis of marginalized communities is imperative to understanding colonial relationships, drawing boundaries between marginalized and dominant communities may exacerbate negative impacts of othering already faced by marginalized communities.

In clustering languages according to their regional origin, Figure 1 omits the fact that languages are adapted and transformed beyond their origins. Furthermore, we rely on Ethnologue for this information, a resource managed by SIL International, for this information. SIL itself is an organization with deep colonial roots (DelValls, 1978). While these combine to paint an incomplete picture, we do so to highlight that the major presence of researchers in nations with indigenous communities, such as the US, Canada, Australia, and Aotearoa (New Zealand), does not lead to representation of the indigenous languages of those nations.

Notably, we diverge from some prior works in one notable aspect, our lack of the use of the word "decolonizing" or "decolonial". We do so not because we believe that coloniality should not be dismantled in our field, but instead because these terms have a tendency to soften tangible requests for land return (Tuck and Yang, 2012).

[7] UN List of Non-Self-Governing Territories

References

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen H. Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Aremu Anuoluwapo, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Rabiu Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. MasakhaNER: Named Entity Recognition for African Languages. *Transactions of the Association for Computational Linguistics*, 9:1116–1131.

David Ifeoluwa Adelani, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende,

Ernie Chang, et al. 2022a. A Few Thousand Translations Go A Long Way! Leveraging Pre-trained Models for African News Translation. *arXiv preprint arXiv:2205.02022*.

David Ifeoluwa Adelani, Graham Neubig, Sebastian Ruder, Shruti Rijhwani, Michael Beukman, Chester Palen-Michel, Constantine Lignos, Jesujoba O Alabi, Shamsuddeen H Muhammad, Peter Nabende, et al. 2022b. MasakhaNER 2.0: Africa-centric Transfer Learning for Named Entity Recognition. *arXiv preprint arXiv:2210.12391*.

Orevaoghene Ahia, Julia Kreutzer, and Sara Hooker. 2021. The Low-resource Double Bind: An Empirical Study of Pruning for Low-resource Machine Translation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3316–3333, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Orevaoghene Ahia, Sachin Kumar, Hila Gonen, Jungo Kasai, David R. Mortensen, Noah A. Smith, and Yulia Tsvetkov. 2023. Do All Languages Cost the Same? Tokenization in the Era of Commercial Language Models.

Alham Fikri Aji, Genta Indra Winata, Fajri Koto, Samuel Cahyawijaya, Ade Romadhony, Rahmad Mahendra, Kemal Kurniawan, David Moeljadi, Radityo Eko Prasojo, Timothy Baldwin, Jey Han Lau, and Sebastian Ruder. 2022. One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7226–7249, Dublin, Ireland. Association for Computational Linguistics.

Idris Akinade, Jesujoba Alabi, David Adelani, Clement Odoje, and Dietrich Klakow. 2023. Varepsilon kú Mask: Integrating Yorùbá Cultural Greetings into Machine Translation. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 1–7, Dubrovnik, Croatia. Association for Computational Linguistics.

Adriana Alvarado Garcia, Juan F Maestre, Manuhuia Barcham, Marilyn Iriarte, Marisol Wong-Villacres, Oscar A Lemus, Palak Dudani, Pedro Reynolds-Cuéllar, Ruotong Wang, and Teresa Cerratto Pargman. 2021. Decolonial Pathways: Our manifesto for a decolonizing Agenda in HCI Research and Design. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–9.

Ashton Anderson, Dan Jurafsky, and Daniel A. McFarland. 2012. Towards a Computational History of the ACL: 1980-2008. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries*, pages 13–21, Jeju Island, Korea. Association for Computational Linguistics.

Arjun Appadurai. 1988. *The Social Life of Things: Commodities in Cultural Perspective*. Cambridge University Press.

Ruth-Ann Armstrong, John Hewitt, and Christopher Manning. 2022. JamPatoisNLI: A Jamaican Patois Natural Language Inference Dataset. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 5307–5320, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Arnav Arora, Lucie-aimée Kaffee, and Isabelle Augenstein. 2023. Probing Pre-Trained Language Models for Cross-Cultural Differences in Values. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 114–130, Dubrovnik, Croatia. Association for Computational Linguistics.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2020a. Translation Artifacts in Cross-lingual Transfer Learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7674–7684, Online. Association for Computational Linguistics.

Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. 2020b. On the Cross-lingual Transferability of Monolingual Representations. In *Proceedings of ACL 2020*.

Mikel Artetxe and Holger Schwenk. 2019. Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond. *Transactions of the ACL 2019*.

Akari Asai, Jungo Kasai, Jonathan Clark, Kenton Lee, Eunsol Choi, and Hannaneh Hajishirzi. 2021. XOR QA: Cross-lingual Open-Retrieval Question Answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 547–564, Online. Association for Computational Linguistics.

François Bar, Matthew S Weber, and Francis Pisani. 2016. Mobile Technology Appropriation in a Distant Mirror: Baroquization, Areolization, and Cannibalism. *New media & society*, 18(4):617–636.

Lisa Bauer, Hanna Tischer, and Mohit Bansal. 2023. Social Commonsense for Explanation and Cultural Bias Discovery. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 3745–3760, Dubrovnik, Croatia. Association for Computational Linguistics.

L. Baumstark, N. Bauer, F. Benke, C. Bertram, S. Bi, C. C. Gong, J. P. Dietrich, A. Dirnaichner, A. Giannousakis, J. Hilaire, D. Klein, J. Koch, M. Leimbach, A. Levesque, S. Madeddu, A. Malik, A. Merfort, L. Merfort, A. Odenweller, M. Pehl, R. C. Pietzcker, F. Piontek, S. Rauner, R. Rodrigues, M. Rottoli, F. Schreyer, A. Schultes, B. Soergel, D. Soergel, J. Strefler, F. Ueckerdt, E. Kriegler, and G. Luderer.

2021. REMIND2.1: Transformation and Innovation Dynamics of the Energy-economic System within Climate and Sustainability Limits. *Geoscientific Model Development*, 14(10):6571–6603.

Derrick A Bell Jr. 1980. Brown v. Board of Education and the Interest-convergence Dilemma. *Harvard law review*, pages 518–533.

BIG bench authors. 2023. Beyond the imitation game: Quantifying and extrapolating the capabilities of language models. *Transactions on Machine Learning Research*.

Emily M Bender. 2011. On Achieving and Evaluating Language-independence in NLP. *Linguistic Issues in Language Technology*, 6.

Emily M. Bender and Batya Friedman. 2018. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. *Transactions of the Association for Computational Linguistics*, 6:587–604.

Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 610–623.

James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. 2010. Theano: A CPU and GPU Math Compiler in Python. In *Proc. 9th python in science conf*, volume 1, pages 3–10.

Steven Bird. 2020. Decolonising Speech and Language Technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519.

Steven Bird. 2022. Local Languages, Third Spaces, and Other High-resource Scenarios. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7817–7829.

Abeba Birhane. 2020. Algorithmic Colonization of Africa. *SCRIPTed*, 17:389.

Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022a. *Power to the People? Opportunities and Challenges for Participatory AI*. Association for Computing Machinery, New York, NY, USA.

Abeba Birhane, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. 2022b. The Values Encoded in Machine Learning Research. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 173–184.

Abeba Birhane and Vinay Uday Prabhu. 2021. Large Image Datasets: A Pyrrhic Win for Computer Vision? In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1536–1546. IEEE.

Claudia Magallanes Blanco, Sabelo Mhlambi, Nanjala Nyabola, Nick Couldry, Toussaint Nothias, and Kathleen Siminyu. 2023. Decolonizing Data, One Language at a Time.

Su Lin Blodgett and Brendan O'Connor. 2017. Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English. *arXiv preprint arXiv:1707.00061*.

Rishi Bommasani. 2023. Evaluation for Change. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8227–8239, Toronto, Canada. Association for Computational Linguistics.

Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the Opportunities and Risks of Foundation Models. *arXiv preprint arXiv:2108.07258*.

Rishi Bommasani, Dilara Soylu, Thomas I Liao, Kathleen A Creel, and Percy Liang. 2023. Ecosystem Graphs: The Social Footprint of Foundation Models. *arXiv preprint arXiv:2303.15772*.

Nelius Boshoff. 2009. Neo-colonialism and Research Collaboration in Central Africa. *Scientometrics*, 81:413–434.

Rodrigo Naumann Boufleur. 2006. *A Questão Da Gambiarra*. Rodrigo Naumann Boufleur.

Geoffrey Bowker and Susan Leigh Star. 1999. Sorting Things Out. *Classification and its consequences*, 4.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language Models are Few-shot [l]earners. *Advances in neural information processing systems*, 33:1877–1901.

Erik Brynjolfsson, Xiang Hui, and Meng Liu. 2019. Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform. *Management Science*, 65(12):5449–5460.

Erik Brynjolfsson and Tom Mitchell. 2017. What can Machine Learning Do? Workforce Implications. *Science*, 358(6370):1530–1534.

Pratik Chakrabarti. 2010. Materials and Medicine: Trade, Conquest and Therapeutics in the Eighteenth Century.

Myra Cheng, Esin Durmus, and Dan Jurafsky. 2023. Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1:*

Long Papers), pages 1504–1532, Toronto, Canada. Association for Computational Linguistics.

Nancy Chinchor. 1991. MUC-3 Evaluation Metrics. In *Third Message Understanding Conference (MUC-3): Proceedings of a Conference Held in San Diego, California, May 21-23, 1991*.

Hyung Won Chung, Xavier Garcia, Adam Roberts, Yi Tay, Orhan Firat, Sharan Narang, and Noah Constant. 2023. UniMax: Fairer and More Effective Language Sampling for Large-Scale Multilingual Pretraining. In *The Eleventh International Conference on Learning Representations*.

Jonathan H. Clark, Eunsol Choi, Michael Collins, Dan Garrette, Tom Kwiatkowski, Vitaly Nikolaev, and Jennimaria Palomaki. 2020. TyDi QA: A Benchmark for Information-Seeking Question Answering in Typologically Diverse Languages. In *Transactions of the Association of Computational Linguistics*.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised Cross-lingual Representation Learning at Scale.

Alexis Conneau, Guillaume Lample, Ruty Rinott, Adina Williams, Samuel R Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. XNLI: Evaluating Cross-lingual Sentence Representations. *arXiv preprint arXiv:1809.05053*.

Jamell Dacon. 2022. Towards a Deep Multi-layered Dialectal Language Analysis: A Case Study of African-American English. *arXiv preprint arXiv:2206.08978*.

Thomas Davidson, Debasmita Bhattacharya, and Ingmar Weber. 2019. Racial Bias in Hate Speech and Abusive Language Detection Datasets. *arXiv preprint arXiv:1905.12516*.

TA DelValls. 1978. El Instituto Lingüístico de Verano, Instrumento del Imperialismo. *Nueva Antropología*, 3(9):117–142.

Alan Dix. 2007. Designing for Appropriation. In *Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK 21*, pages 1–4.

Quinn Dombrowski. 2020. Preparing Non-english Texts for Computational Analysis. *Modern Languages Open*.

Rachel Dorn. 2019. Dialect-Specific Models for Automatic Speech Recognition of African American Vernacular English. In *Proceedings of the Student Research Workshop Associated with RANLP 2019*, pages 16–20, Varna, Bulgaria. INCOMA Ltd.

Esin Durmus, Karina Nyugen, Thomas I. Liao, Nicholas Schiefer, Amanda Askell, Anton Bakhtin, Carol Chen, Zac Hatfield-Dodds, Danny Hernandez, Nicholas Joseph, Liane Lovitt, Sam McCandlish, Orowa Sikder, Alex Tamkin, Janel Thamkul, Jared Kaplan, Jack Clark, and Deep Ganguli. 2023. Towards Measuring the Representation of Subjective Global Opinions in Language Models.

Abteen Ebrahimi, Manuel Mager, Arturo Oncevay, Vishrav Chaudhary, Luis Chiruzzo, Angela Fan, John Ortega, Ricardo Ramos, Annette Rios, Ivan Vladimir Meza Ruiz, Gustavo Giménez-Lugo, Elisabeth Mager, Graham Neubig, Alexis Palmer, Rolando Coto-Solano, Thang Vu, and Katharina Kann. 2022. AmericasNLI: Evaluating Zero-shot Natural Language Understanding of Pretrained Multilingual Models in Truly Low-resource Languages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6279–6299, Dublin, Ireland. Association for Computational Linguistics.

Abteen Ebrahimi, Manuel Mager, Shruti Rijhwani, Enora Rice, Arturo Oncevay, Claudia Baltazar, María Cortés, Cynthia Montaño, John E. Ortega, Rolando Coto-solano, Hilaria Cruz, Alexis Palmer, and Katharina Kann. 2023. Findings of the AmericasNLP 2023 Shared Task on Machine Translation into Indigenous Languages. In *Proceedings of the Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP)*, pages 206–219, Toronto, Canada. Association for Computational Linguistics.

Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *arXiv preprint arXiv:2303.10130*.

Joseph Errington. 2001. Colonial Linguistics. *Annual Review of Anthropology*, 30(1):19–39.

Arturo Escobar. 2011. Sustainability: Design for the Pluriverse. *Development*, 54:137–140.

Kawin Ethayarajh and Dan Jurafsky. 2020. Utility is in the eye of the user: A critique of NLP leaderboards. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4846–4853, Online. Association for Computational Linguistics.

Fahim Faisal, Yinkai Wang, and Antonios Anastasopoulos. 2022. Dataset Geography: Mapping Language Data to Language Users. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3381–3411, Dublin, Ireland. Association for Computational Linguistics.

Frantz Fanon. 1959. *A Dying Colonialism*.

Ali Farghaly and Khaled Shaalan. 2009. Arabic Natural Language Processing: Challenges and Solutions. *ACM Transactions on Asian Language Information Processing (TALIP)*, 8(4):1–22.

Malcom Ferdinand. 2021. *Decolonial Ecology: Thinking from the Caribbean World*. John Wiley & Sons.

Jack FitzGerald, Christopher Hench, Charith Peris, Scott Mackie, Kay Rottmann, Ana Sanchez, Aaron Nash, Liam Urbach, Vishesh Kakarala, Richa Singh, Swetha Ranganath, Laurie Crist, Misha Britan, Wouter Leeuwis, Gokhan Tur, and Prem Natarajan. 2022. MASSIVE: A 1m-Example Multilingual Natural Language Understanding Dataset with 51 Typologically-Diverse Languages.

Christopher Frauenberger. 2019. Entanglement HCI the Next Wave? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(1):1–27.

Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al. 2020. The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *arXiv preprint arXiv:2101.00027*.

DW Gotterbarn, Bo Brinkman, Catherine Flick, Michael S Kirkpatrick, Keith Miller, Kate Vazansky, and Marty J Wolf. 2018. ACM Code of Ethics and Professional Conduct.

Sophie Groenwold, Lily Ou, Aesha Parekh, Samhita Honnavalli, Sharon Levy, Diba Mirza, and William Yang Wang. 2020a. Investigating African-American Vernacular English in Transformer-Based Text Generation. *arXiv preprint arXiv:2010.02510*.

Sophie Groenwold, Lily Ou, Aesha Parekh, Samhita Honnavalli, Sharon Levy, Diba Mirza, and William Yang Wang. 2020b. Investigating African-American Vernacular English in Transformer-Based Text Generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5877–5883, Online. Association for Computational Linguistics.

Ramón A. Gutiérrez. 2004. Internal Colonialism: An American Theory of Race. *Du Bois Review: Social Science Research on Race*, 1(2):281–295.

Oliver L Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–35.

Matan Halevy, Camille Harris, Amy Bruckman, Diyi Yang, and Ayanna Howard. 2021. Mitigating Racial Biases in Toxic Language Detection with an Equity-Based Ensemble Framework. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–11.

Eszter Hargittai. 2001. Second-level Digital Divide: Mapping Differences in People's Online Skills. *arXiv preprint cs/0109068*.

Camille Harris, Matan Halevy, Ayanna Howard, Amy Bruckman, and Diyi Yang. 2022. Exploring the Role of Grammar and Word Choice in Bias Toward African American English (AAE) in Hate Speech Classification. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 789–798.

Camille Harris, Amber Gayle Johnson, Sadie Palmer, Diyi Yang, and Amy Bruckman. 2023. "Honestly, I Think TikTok Has a Vendetta Against Black Creators": Understanding Black Content Creator Experiences on TikTok. 7(CSCW2).

Charles T. Hemphill, John J. Godfrey, and George R. Doddington. 1990. The ATIS Spoken Language Systems Pilot Corpus. In *Speech and Natural Language: Proceedings of a Workshop Held at Hidden Valley, Pennsylvania, June 24-27,1990*.

Peter Henderson, Jieru Hu, Joshua Romoff, Emma Brunskill, Dan Jurafsky, and Joelle Pineau. 2020. Towards the Systematic Reporting of the Energy and Carbon Footprints of Machine Learning. *J. Mach. Learn. Res.*, 21(1).

Linette N Hinton and Karen E Pollock. 2000. Regional Variations in the Phonological Characteristics of African American Vernacular English. *World Englishes*, 19(1):59–71.

Julia Hirschberg. 1998. Every Time I Fire a Linguist, my Performance Goes Up, and Other Myths of the Statistical natural Language Processing Revolution. In *Fifteenth National Conference on Artificial Intelligence (AAAI-98)*.

Ian Hodder. 2012. Entangled: An Archaeology of the Relationships between Humans and Things.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, Tom Hennigan, Eric Noland, Katie Millican, George van den Driessche, Bogdan Damoc, Aurelia Guy, Simon Osindero, Karen Simonyan, Erich Elsen, Jack W. Rae, Oriol Vinyals, and Laurent Sifre. 2022. Training Compute-Optimal Large Language Models.

Sara Hooker. 2021. The Hardware Lottery. *Communications of the ACM*, 64(12):58–65.

Dirk Hovy and Shannon L Spruit. 2016. The Social Impact of Natural Language Processing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 591–598.

Dirk Hovy and Diyi Yang. 2021. The Importance of Modeling Social Factors of Language: Theory and Practice. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 588–602, Online. Association for Computational Linguistics.

Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. 2020. XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalisation. In *Proceedings of the 37th International*

Conference on Machine Learning, volume 119 of *Proceedings of Machine Learning Research*, pages 4411–4421. PMLR.

Jacquelyne Johnson Jackson. 1997. On Oakland's Ebonics. *The Black Scholar*, 27(1):18–25.

Rebecca L Johnson, Giada Pistilli, Natalia Menédez-González, Leslye Denisse Dias Duran, Enrico Panai, Julija Kalpokiene, and Donald Jay Bertulfo. 2022. The Ghost in the Machine has an American Accent: Value Conflict in GPT-3.

Anna Jørgensen, Dirk Hovy, Anders Søgaard, et al. 2016. Learning a POS Tagger for AAVE-like Language. In *The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Proceedings of the conference*. Association for Computational Linguistics.

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The State and Fate of Linguistic Diversity and Inclusion in the NLP World. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.

David Jurgens, Yulia Tsvetkov, and Dan Jurafsky. 2017a. Incorporating Dialectal Variability for Socially Equitable Language Identification. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 51–57.

David Jurgens, Yulia Tsvetkov, and Dan Jurafsky. 2017b. Writer Profiling without the Writer's Text. In *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II 9*, pages 537–558. Springer.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.

Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. 2020. Scaling Laws for Neural Language Models.

Webb Keane. 2003. Semiotics and the Social Analysis of Material Things. *Language & communication*, 23(3-4):409–425.

Sharese King. 2020. From African American Vernacular English to African American Language: Rethinking the Study of Race and Language in African Americans' Speech. *Annual Review of Linguistics*, 6:285–300.

Dorothea Kleine. 2009. ICT4WHAT?-Using the Choice Framework to Operationalise the Capability Approach to Development. In *2009 international conference on information and communication technologies and development (ICTD)*, pages 108–117. IEEE.

Maciej K Konopiński. 2020. Shannon Diversity Index: a Call to Replace the Original Shannon's Formula with Unbiased Estimator in the Population Genetics Studies. *PeerJ*, 8:e9391.

András Kornai. 2013. Digital Language Death. *PloS one*, 8(10):e77056.

Julia Kreutzer, Isaac Caswell, Lisa Wang, Ahsan Wahab, Daan van Esch, Nasanbayar Ulzii-Orshikh, Allahsera Tapo, Nishant Subramani, Artem Sokolov, Claytone Sikasote, Monang Setyawan, Supheakmungkol Sarin, Sokhar Samb, Benoît Sagot, Clara Rivera, Annette Rios, Isabel Papadimitriou, Salomey Osei, Pedro Ortiz Suarez, Iroro Orife, Kelechi Ogueji, Andre Niyongabo Rubungo, Toan Q. Nguyen, Mathias Müller, André Müller, Shamsuddeen Hassan Muhammad, Nanda Muhammad, Ayanda Mnyakeni, Jamshidbek Mirzakhalov, Tapiwanashe Matangira, Colin Leong, Nze Lawson, Sneha Kudugunta, Yacine Jernite, Mathias Jenny, Orhan Firat, Bonaventure F. P. Dossou, Sakhile Dlamini, Nisansa de Silva, Sakine Çabuk Ballı, Stella Biderman, Alessia Battisti, Ahmed Baruwa, Ankur Bapna, Pallavi Baljekar, Israel Abebe Azime, Ayodele Awokoya, Duygu Ataman, Orevaoghene Ahia, Oghenefego Ahia, Sweta Agrawal, and Mofetoluwa Adeyemi. 2022. Quality at a Glance: An Audit of Web-Crawled Multilingual Datasets. *Transactions of the Association for Computational Linguistics*, 10:50–72.

Neha Kumar and Nimmi Rangaswamy. 2013. The Mobile Media Actor-Network in Urban India. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1989–1998.

Bruno Latour. 2005. Reassembling the Social: an Introduction to Actor-Network-Theory. *Oxford: Clarendon, 2005*.

Nayeon Lee, Chani Jung, and Alice Oh. 2023. Hate Speech Classifiers are Culturally Insensitive. In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 35–46, Dubrovnik, Croatia. Association for Computational Linguistics.

Heather Lent, Emanuele Bugliarello, Miryam de Lhoneux, Chen Qiu, and Anders Søgaard. 2021. On Language Models for Creoles. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 58–71, Online. Association for Computational Linguistics.

Heather Lent, Kelechi Ogueji, Miryam de Lhoneux, Orevaoghene Ahia, and Anders Søgaard. 2022. What a Creole Wants, What a Creole Needs. *arXiv preprint arXiv:2206.00437*.

Jacqueline Léon. 2021. *Automating Linguistics*. Springer.

Patrick Lewis, Barlas Oğuz, Ruty Rinott, Sebastian Riedel, and Holger Schwenk. 2020. MLQA: Evaluating Cross-lingual Extractive Question Answering. In *Proceedings of ACL 2020*.

Haoran Li, Abhinav Arora, Shuohui Chen, Anchit Gupta, Sonal Gupta, and Yashar Mehdad. 2021. MTOP: A Comprehensive Multilingual Task-Oriented Semantic Parsing Benchmark. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2950–2962, Online. Association for Computational Linguistics.

Max Liboiron. 2021. *Pollution is Colonialism*. Duke University Press.

Shayne Longpre, Robert Mahari, Anthony Chen, Naana Obeng-Marnu, Damien Sileo, William Brannon, Niklas Muennighoff, Nathan Khazam, Jad Kabbara, Kartik Perisetla, Xinyi Wu, Enrico Shippole, Kurt Bollacker, Tongshuang Wu, Luis Villa, Sandy Pentland, Deb Roy, and Sara Hooker. 2023. The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI.

Patrice Lopez. 2009. GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications. In *Research and Advanced Technology for Digital Libraries: 13th European Conference, ECDL 2009, Corfu, Greece, September 27-October 2, 2009. Proceedings 13*, pages 473–474. Springer.

Anastassia Loukina, Nitin Madnani, and Klaus Zechner. 2019. The Many Dimensions of Algorithmic Fairness in Educational Applications. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 1–10.

Malte Luebker. 2010. Inequality, Income Shares and Poverty: The Practical Meaning of Gini Coefficients. Technical report, International Labour Organization.

Manuel Mager, Ximena Gutierrez-Vasques, Gerardo Sierra, and Ivan Meza-Ruiz. 2018. Challenges of Language Technologies for the Indigenous Languages of the Americas. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 55–69, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Manuel Mager, Elisabeth Mager, Katharina Kann, and Ngoc Thang Vu. 2023. Ethical Considerations for Machine Translation of Indigenous Languages: Giving a Voice to the Speakers. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4871–4897, Toronto, Canada. Association for Computational Linguistics.

Manuel Mager, Arturo Oncevay, Abteen Ebrahimi, John Ortega, Annette Rios, Angela Fan, Ximena Gutierrez-Vasques, Luis Chiruzzo, Gustavo Giménez-Lugo, Ricardo Ramos, Ivan Vladimir Meza Ruiz, Rolando Coto-Solano, Alexis Palmer, Elisabeth Mager-Hois, Vishrav Chaudhary, Graham Neubig, Ngoc Thang Vu, and Katharina Kann. 2021. Findings of the AmericasNLP 2021 Shared Task on Open Machine Translation for Indigenous Languages of the Americas. In *Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas*, pages 202–217, Online. Association for Computational Linguistics.

Anne E Magurran. 2021. Measuring Biological Diversity. *Current Biology*, 31(19):R1174–R1177.

Nelson Maldonado-Torres. 2007. On the Coloniality of Being: Contributions to the Development of a Concept. *Cultural studies*, 21(2-3):240–270.

Christopher Manning and Hinrich Schutze. 1999. *Foundations of Statistical Natural Language Processing*. MIT press.

Zion Mengesha, Courtney Heldreth, Michal Lahav, Juliana Sublewski, and Elyse Tuennerman. 2021. "i don't think these devices are very culturally sensitive."—Impact of Automated Speech Recognition Errors on African Americans. *Frontiers in Artificial Intelligence*, 4:169.

Arnav Mhaske, Harshit Kedia, Sumanth Doddapaneni, Mitesh M. Khapra, Pratyush Kumar, Rudra Murthy, and Anoop Kunchukuttan. 2023. Naamapadam: A Large-Scale Named Entity Annotated Data for Indic Languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10441–10456, Toronto, Canada. Association for Computational Linguistics.

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 220–229.

Shakir Mohamed, Marie-Therese Png, and William Isaac. 2020. Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33:659–684.

Saif M Mohammad, Mohammad Salameh, and Svetlana Kiritchenko. 2016. How Translation Alters Sentiment. *Journal of Artificial Intelligence Research*, 55:95–130.

Thomas S Mullaney. 2017. *The Chinese Typewriter: A History*. MIT Press.

Tarek Naous, Michael J Ryan, and Wei Xu. 2023. Having Beer after Prayer? Measuring Cultural Bias in Large Language Models. *arXiv preprint arXiv:2305.14456*.

Nature. 2014. Introducing the Index. *Nature*, 515:S52–S53.

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. 2020. Participatory Research for Low-resourced Machine Translation: A Case Study in African Languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2144–2160, Online. Association for Computational Linguistics.

Joakim Nivre, Mitchell Abrams, Željko Agić, Lars Ahrenberg, Lene Antonsen, Maria Jesus Aranzabe, Gashaw Arutie, Masayuki Asahara, Luma Ateyah, Mohammed Attia, et al. 2018. Universal Dependencies 2.2.

Daniel Nkemelu, Harshil Shah, Michael Best, and Irfan Essa. 2022. Tackling Hate Speech in Low-resource Languages with Context Experts. In *Proceedings of the 2022 International Conference on Information and Communication Technologies and Development*, pages 1–11.

Ihudiya Finda Ogbonnaya-Ogburu, Angela DR Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–16.

Tolúlọpẹ́ Ògúnrẹ̀mí, Dan Jurafsky, and Christopher Manning. 2023. Mini But Mighty: Efficient Multilingual Pretraining with Linguistically-informed Data Selection. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1251–1266, Dubrovnik, Croatia. Association for Computational Linguistics.

Wilhelmina Onyothi Ògúnrẹ̀mí, Tolúlọpẹ́and Nekoto and Saron Samuel. 2023. Decolonizing NLP for "Low-resource Languages": Applying Abeba Birhane's Relational Ethics. *GRACE: Global Review of AI Community Ethics*, 1(1).

Iroro Orife, Julia Kreutzer, Blessing Sibanda, Daniel Whitenack, Kathleen Siminyu, Laura Martinus, Jamiil Toure Ali, Jade Abbott, Vukosi Marivate, Salomon Kabongo, et al. 2020. Masakhane–Machine Translation for Africa. *arXiv preprint arXiv:2003.11529*.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training Language Models to Follow Instructions with Human Feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual Name Tagging and Linking for 282 Languages. In *Proceedings of ACL 2017*, pages 1946–1958.

John Paolillo, Daniel Pimienta, Daniel Prado, et al. 2005. Measuring Linguistic Diversity on the Internet. *UNESCO Institute for Statistics Montreal, Canada*.

Isabel Papadimitriou, Kezia Lopez, and Dan Jurafsky. 2023. Multilingual BERT has an Accent: Evaluating English Influences on Fluency in Multilingual Models. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1194–1200, Dubrovnik, Croatia. Association for Computational Linguistics.

Patrick Paroubek, Stéphane Chaudiron, and Lynette Hirschman. 2007. Principles of Evaluation in Natural Language Processing. *Revue TAL*, 48(1):7–31.

David Patterson, Joseph Gonzalez, Quoc Le, Chen Liang, Lluis-Miquel Munguia, Daniel Rothchild, David So, Maud Texier, and Jeff Dean. 2021. Carbon Emissions and Large Neural Network Training. *arXiv preprint arXiv:2104.10350*.

Daniel Pimienta, Daniel Prado, and Álvaro Blanco. 2009. Twelve Years of Measuring Linguistic Diversity in the Internet: Balance and Perspectives.

Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How Multilingual is Multilingual BERT? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy. Association for Computational Linguistics.

Edoardo Maria Ponti, Goran Glavaš, Olga Majewska, Qianchu Liu, Ivan Vulić, and Anna Korhonen. 2020. XCOPA: A Multilingual Dataset for Causal Commonsense Reasoning.

Anibal Quijano. 2000. Coloniality of Power and Eurocentrism in Latin America. *International sociology*, 15(2):215–232.

Inioluwa Deborah Raji, Emily M Bender, Amandalynne Paullada, Emily Denton, and Alex Hanna. 2021. AI and the Everything in the Whole Wide World Benchmark. *arXiv preprint arXiv:2111.15366*.

Nimmi Rangaswamy and Melissa Densmore. 2013. Understanding Jugaad: ICTD and the Tensions of Appropriation, Innovation and Utility. In *Proceedings*

of the Sixth International Conference on Information and Communications Technologies and Development: Notes-Volume 2, pages 120–123.

Shruti Rijhwani, Antonios Anastasopoulos, and Graham Neubig. 2020. OCR Post Correction for Endangered Language Texts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5931–5942, Online. Association for Computational Linguistics.

Shaurya Rohatgi, Yanxia Qin, Benjamin Aw, Niranjana Unnithan, and Min-Yen Kan. 2023. The ACL OCL Corpus: Advancing Open Science in Computational Linguistics. *arXiv preprint arXiv:2305.14996*.

Maggie Ronkin and Helen E Karn. 1999. Mock Ebonics: Linguistic Racism in Parodies of Ebonics on the Internet. *Journal of sociolinguistics*, 3(3):360–380.

Sebastian Ruder, Noah Constant, Jan Botha, Aditya Siddhant, Orhan Firat, Jinlan Fu, Pengfei Liu, Junjie Hu, Dan Garrette, Graham Neubig, et al. 2021. XTREME-R: Towards More Challenging and Nuanced Multilingual Evaluation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10215–10245.

Sebastian Ruder, Ivan Vulić, and Anders Søgaard. 2022. Square One Bias in NLP: Towards a Multi-Dimensional Exploration of the Research Manifold. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2340–2354, Dublin, Ireland. Association for Computational Linguistics.

Mukund Rungta, Janvijay Singh, Saif M Mohammad, and Diyi Yang. 2022. Geographic Citation Gaps in NLP Research. *arXiv preprint arXiv:2210.14424*.

Edward Said. 1978. Orientalism.

Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–15.

Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. 2019. The Risk of Racial Bias in Hate Speech Detection. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 1668–1678.

Naomi Saphra, Eve Fleisig, Kyunghyun Cho, and Adam Lopez. 2023. First Tragedy, then Parse: History Repeats Itself in the New Era of Large Language Models.

Lane Schwartz. 2022. Primum Non Nocere: Before working with Indigenous data, the ACL must confront ongoing colonialism. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 724–731, Dublin, Ireland. Association for Computational Linguistics.

Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 59–68.

Amartya Sen. 1995. *Inequality Reexamined*. Harvard University Press.

Amartya Sen. 2001. *Development as Freedom*. Oxford Paperbacks.

Jaime Sevilla, Lennart Heim, Anson Ho, Tamay Besiroglu, Marius Hobbhahn, and Pablo Villalobos. 2022. Compute Trends Across Three Eras of Machine Learning.

Omar Shaikh, Hongxin Zhang, William Held, Michael Bernstein, and Diyi Yang. 2023. On second thought, let's not think step by step! bias and toxicity in zero-shot reasoning. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4454–4470, Toronto, Canada. Association for Computational Linguistics.

Andrew L Sheldon. 1969. Equitability Indices: Dependence on the Species Count. *Ecology*, 50(3):466–467.

Boaz Shmueli, Jan Fell, Soumya Ray, and Lun-Wei Ku. 2021. Beyond Fair Pay: Ethical Implications of NLP Crowdsourcing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3758–3769, Online. Association for Computational Linguistics.

Laura Smith, Salvatore Giorgi, Rishi Solanki, Johannes Eichstaedt, H. Andrew Schwartz, Muhammad Abdul-Mageed, Anneke Buffone, and Lyle Ungar. 2016. Does 'well-being' Translate on Twitter? In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2042–2047, Austin, Texas. Association for Computational Linguistics.

Linda Tuhiwai Smith. 1999. Decolonizing Methodologies: Research and Indigenous Peoples.

Hanna L Smokoski. 2016. Voicing the Other: Mock AAVE on Social Media.

Ramesh Srinivasan. 2006. Indigenous, Ethnic and Cultural Articulations of New Media. *International journal of cultural studies*, 9(4):497–518.

Carolyne Stanforth. 2006. Using Actor-Network Theory to Analyze e-Government Implementation in Developing Countries. *Information Technologies & International Development*, 3(3):pp–35.

Stephanie Strassel and Andrew W Cole. 2006. Corpus Development and Publication. *Proceedings of LREC, Genoa, Italy. http://papers. ldc. upenn. edu/LREC2006/Co rpusDevelopmentAndPublication. pdf.*

Emma Strubell, Ananya Ganesh, and Andrew McCallum. 2019. Energy and Policy Considerations for Deep Learning in NLP. *arXiv preprint arXiv:1906.02243*.

Beth M. Sundheim. 1995. Overview of Results of the MUC-6 Evaluation. In *Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995*.

Divy Hasmukhbhai Thakkar, Azra Ismail, Pratyush Kumar, Alex Hanna, Nithya Sambasivan, and Neha Kumar. 2022. When is ML Data Good?: Valuing in Public Health Datafication.

Zoe Todd. 2016. An Indigenous Feminist's Take on the Ontological Turn:'Ontology'is Just Another Word for Colonialism. *Journal of historical sociology*, 29(1):4–22.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open Foundation and Fine-tuned Chat Models. *arXiv preprint arXiv:2307.09288*.

Eve Tuck and K Wayne Yang. 2012. Decolonization is not a Metaphor. *Education & Society*, 1(1):1–40.

Ngugi Wa Thiong'o. 1992. *Decolonising the Mind: The Politics of Language in African Literature*. East African Publishers.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R Bowman. 2018. GLUE: A Multi-task Benchmark and Analysis Platform for Natural Language Understanding. *arXiv preprint arXiv:1804.07461*.

Rose Wang and Dorottya Demszky. 2023. Is ChatGPT a Good Teacher Coach? Measuring Zero-Shot Performance For Scoring and Providing Actionable Insights on Classroom Instruction. In *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, pages 626–667, Toronto, Canada. Association for Computational Linguistics.

Alex Warstadt, Leshem Choshen, Aaron Mueller, Adina Williams, Ethan Wilcox, and Chengxu Zhuang. 2023. The BabyLM Challenge: Sample-efficient Pretraining on a Developmentally Plausible Corpus.

Laura Weidinger, Jonathan Uesato, Maribeth Rauh, Conor Griffin, Po-Sen Huang, John Mellor, Amelia Glaese, Myra Cheng, Borja Balle, Atoosa Kasirzadeh, et al. 2022. Taxonomy of Risks Posed by Language Models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 214–229.

Lee Wengraf. 2018. *Extracting Profit: Imperialism, Neoliberalism and the New Scramble for Africa*. Haymarket Books.

Donald Winford. 2015. The Origins of African American Vernacular English. *The Oxford handbook of African American language*, page 85.

Langdon Winner. 1980. Do Artifacts Have Politics? *Daedalus*, 109(1):121–136.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.

Yinfei Yang, Yuan Zhang, Chris Tar, and Jason Baldridge. 2019. PAWS-X: A Cross-lingual Adversarial Dataset for Paraphrase Identification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3687–3692, Hong Kong, China. Association for Computational Linguistics.

Zheng-Xin Yong, Cristina Menghini, and Stephen H. Bach. 2023. Low-Resource Languages Jailbreak GPT-4.

Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic Lies: Understanding AI-generated Misinformation and Evaluating Algorithmic and Human Solutions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–20.

Caleb Ziems, Jiaao Chen, Camille Harris, Jessica Anderson, and Diyi Yang. 2022. VALUE: Understanding Dialect Disparity in NLU. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3701–3720.

Caleb Ziems, William Held, Jingfeng Yang, Jwala Dhamala, Rahul Gupta, and Diyi Yang. 2023. Multi-VALUE: A framework for cross-dialectal English NLP. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 744–768, Toronto, Canada. Association for Computational Linguistics.

Simon Zwarts and Mark Dras. 2007. Statistical Machine Translation of Australian Aboriginal Languages: Morphological Analysis with Languages of Differing Morphological Richness. In *Proceedings of the Australasian Language Technology Workshop 2007*, pages 134–142.

Pierre Zweigenbaum, Serge Sharoff, and Reinhard Rapp. 2018. Spotting Parallel Sentences in Comparable Corpora. In *Proceedings of 11th Workshop on Building and Using Comparable Corpora*, pages 39–42.

A Author Affiliation Methodology

In an ideal world, each paper would be mapped to a location by the authors themselves. This was often the practice prior to the prevalence of email when mailing addresses were more relevant to academic communication. However, in practice, authors often report different aspects of information in each paper. We use the following cascade of rules to map each author on a paper to their affiliation:

1. If the author lists an email address with a country-level top-level domain, e.g. "cn" for China, we assign the affiliation as that country.

2. If the author lists a country name explicitly, we assign the affiliation as that country.

3. If the author lists a postal code, we get all possible cities associated with the postalcode from GeoNames. If the author also lists exactly one of these cities, we assign the affiliation as the country which contains the matched city and postal code pair.

4. If the author lists an University email address with a domain in the list from Rungta et al. (2022), we assign the affiliation as the country associated with that University.

5. If the author lists an institution name which has at most one character difference from an institution name in the list from Rungta et al. (2022), we assign the affiliation as the country associated with that University.

We evaluate these rules in the order listed above, prioritizing rules which are less likely to be affected by noise from the GrobID (Lopez, 2009) parsing tool or to incorrectly map affiliations based on text matching alone.

B Language Mention Methodology

Following Joshi et al. (2020), we first search the plain text of all papers in our corpus for mentions of each language listed in their lang2tax file. Upon manual inspection, we find a reasonable number of spurious matches such as homonyms of the language, *"...how they work (Serrano and Smith, 2019..."* being counted for the Serrano language, or incidental mentions of languages, "Unlike English..." being counted for English.

To handle this, we filter all mentions using GPT-3.5-Turbo as a filtering tool. We sample 5 sentences which include a full-text language match and prompt the LLM to produce its own set of languages which are the main focus.

```
{
     "role": "system",
     "content": "You are a Natural Language
         Processing expert carefully studying
         papers from ACL. On each line, only
         return valid Python set.",
},
{
     "role": "user",
     "content": str(
  'What are the primary languages of interest of
       a paper with these sentences? Ignore
      languages that are only mentioned in
      passing, for example mentions like "Unlike
       English": should not lead to English
      being included in the set. \n Sentences: '
  + "\n".join(LANG_MENTIONS)])
     ),
},
```

To avoid hallucinations from the language model, we use the intersection between the original set of languages explicitly mentioned in the text and the set of languages returned by GPT-3.5-Turbo. This means that the precision is, at worst, the same as the method from Joshi et al. (2020). Notably, this likely undercounts the concentration of research in English, as many papers may focus on English without following the so-called "Bender Rule" of stating the language of interest regardless (Bender, 2011).

C Paper Categorization Methdology

When developing heuristics, we prioritized precision over recall and manually sampled 30 papers from each group to confirm that all assignments were reasonable by the first authors judgement.

Our heuristics are as follows:

1. All papers from the Language Resources and Evaluation Conference (LREC) are used as a sample of Annotation and Benchmark creation. This results in 2447 papers which have a parseable PDF as our sample. While annotated resources are released via other venues, LREC is both a top venue and almost exclusively focuses on annotated resources and evaluation allowing us to keep with our principle of precision.

2. All papers which use the phrase "our model", "our algorithm", or "our method" in the abstract of the work are used as a sample of Modeling and Methods papers. This results in 4940 papers which have a parseable PDF as our sample.

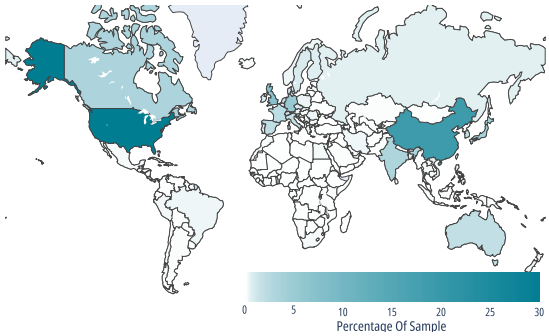


Figure C.1: Geographic density of papers which mention using NVIDIA GPUs. Only includes papers after the ACL added the Responsible NLP Checklist, which asks that all papers to report hardware used.

3. Deployment is the phase which the ACL corpus is least representative. While many companies produce research papers, it is unclear which of these papers are ultimately deployed in products. We opt in favor of precision, counting only papers from workshops and or submission tracks with titles that include "Industry" or "In Practice". This results in 97 papers, which is a relatively small sample compared to all NLP deployment globally.

4. Finally, to get a sense of hardware use in NLP, we collect all papers which mention "NVIDIA", "RTX", or "GeForce". This results in 2663 papers.