Diversity and language technology: how language modeling bias causes epistemic injustice

Paula Helm, Gábor Bella, Gertraud Koch, Fausto Giunchiglia

Ethics and Information Technology (2024) 26:8
https://doi.org/10.1007/s10676-023-09742-6

ORIGINAL PAPER

Diversity and language technology: how language modeling bias causes epistemic injustice

Paula Helm[1] · **Gábor Bella**[2] · **Gertraud Koch**[3] · **Fausto Giunchiglia**[4]

Accepted: 13 December 2023

Abstract

It is well known that AI-based language technology—large language models, machine translation systems, multilingual dictionaries, and corpora—is currently limited to three percent of the world's most widely spoken, financially and politically backed languages. In response, recent efforts have sought to address the "digital language divide" by extending the reach of large language models to "underserved languages." We show how some of these efforts tend to produce flawed solutions that adhere to a hard-wired representational preference for certain languages, which we call language modeling bias. Language modeling bias is a specific and under-studied form of linguistic bias were language technology by design favors certain languages, dialects, or sociolects with respect to others. We show that language modeling bias can result in systems that, while being precise regarding languages and cultures of dominant powers, are limited in the expression of socio-culturally relevant notions of other communities. We further argue that at the root of this problem lies a systematic tendency of technology developer communities to apply a simplistic understanding of diversity which does not do justice to the more profound differences that languages, and ultimately the communities that speak them, embody. Drawing on the concept of epistemic injustice, we point to the broader ethico-political implications and show how it can lead not only to a disregard for valuable aspects of diversity but also to an under-representation of the needs of marginalized language communities. Finally, we present an alternative socio-technical approach that is designed to tackle some of the analyzed problems.

Keywords Language technology · Diversity · Digital divide · Epistemic injustice · Linguistic bias · Language modeling bias · Lexical gaps · Large language models

Introduction

At the latest since the release of products such as DeepL or ChatGPT, AI-supported language technologies are well on their way to becoming mainstream and thus an integral part of everyday communication and work routines. As such, they shape social relationships and influence processes of knowledge production and proliferation. Following science and technology scholar Langdon Winner, language technologies can be defined as inherently political because they warrant processes of profound social change (Winner, 1988). Given that language technologies are both sociotechnical and inherently political, it is important to ask how they privilege certain points of view and how their specific design is influenced by the interests and idea(l)s of certain groups of people. From an ethical point of view, LT applications thus require appropriate reflection of their inherent biases to prevent discriminatory consequences for marginalized groups of people.

Paula Helm and Gábor Bella have contributed equally to this paper.

✉ Paula Helm
p.m.helm@uva.nl

Gábor Bella
gabor.bella@imt-atlantique.fr

Gertraud Koch
gertraud.koch@uni-hamburg.de

Fausto Giunchiglia
fausto.giunchiglia@unitn.it

[1] University of Amsterdam, Amsterdam, The Netherlands

[2] Lab-STICC CNRS UMR 628, IMT Atlantique, Brest, France

[3] University of Hamburg, Hamburg, Germany

[4] University of Trento, Trento, Italy

Published online: 27 January 2024

When it comes to questions of linguistic bias in NLP, Hovy and Prabhumoye (2021) identify five sources of bias: (1) the data,(2) the annotation process, (3) the models, (4) the input representation process and (5) the research design for studying bias. Following up on this, Blodgett et al. (2020) develop three recommendations for further research: (a) better consideration of social hierarchies and the language ideologies created or transported through the systems, (b) the specification of the normative dimensions applied in the analysis with respect to what is harmful and what beneficial when systems are applied, and (c) value-sensitive and community oriented perspectives of NLP systems in use in language practice. While we take these recommendations as a starting point, we also acknowledge two limitations: one is the focus on linguistic dimensions, which does not pay sufficient attention to the biases in the design of the respective technologies and the methodologies behind them, and the other is the lack of a more explicit discussion of the ethico-political dimensions of the problem.

With this paper, we aim to take a first step towards filling these two gaps by extending the understanding of bias in NLP to another dimension of bias at the technological level: hard-wired, but mostly unintentional by-design preferences for certain languages. We further point out the discriminatory consequences caused by what we term *language modeling bias*. These are, we admit, difficult to pin down, perhaps even more so than is the case with racial or sexist biases in NLP (Bender et al., 2021). This is because the type of bias we are pointing at unfolds its problematic effects not primarily at the individual but at the systemic level. Nevertheless, it is important to pay attention to it because it has far-reaching effects on the equal opportunities of different language communities in terms of self-representation, epistemic self-determination, and communicative participation (Nyabola, 2018).

Apart from the research around linguistic bias in NLP, recent debates on what has been coined *digital language divide* are also central when dealing with language modeling bias (Zaug et al., 2022; Young, 2015). Digital language divide refers to the gap between languages with and without a considerable representation within the worldwide digital infrastructure. As shown by Kornai (2013) about 10 years ago, less than 5% of the world's 7–8000 languages have a remotely significant representation on the Internet, and despite the progresses of a decade, the gap has barely shrunk (Joshi et al., 2020). The political dimension of the divide is most evident when reconstructing the argument of size, which of course matters in the rapid upscaling of digital support for certain languages, but is at once a result of imperialist politics and far from the only determining factor in digital support. Consider, for example, that the number of Wikipedia pages for Kiswahili, one of the major African languages with about 80 million speakers, is as high as for Breton, an endangered Celtic language in western France with about 200 thousand speakers (according to optimistic estimates).[1] The former, although widely spoken receives little support and if so, mostly top-down, while the latter benefits from culture preservation programs.

For many members of language communities such as Kiswahili, digital representation is an urgent project (Nyabola, 2018). Following this demand, in the field of language technology, riding the wave of the recent breakthrough of neural AI, the last decade saw a surge in multilingual language tools and resources for 'under-resourced languages.' Researchers have tried to enable technologies such as machine translation, natural language processing, or speech recognition, to an ever larger scope of languages. However, many such efforts are based on a vision according to which, with the help of AI, already successfully developed and applied methods and systems that are designed and sought of from an anglo-centric culture of technology development, are one-to-one adopted to other contexts (Bird, 2020; Schwartz, 2022). This approach to bridging the divide leads to a misalignment between the interests and solutions of the former and the living realities of the latter (Helm et al., 2023). Worse, due to general ignorance of the more profound dimensions of linguistic diversity and ultimately the cultural differences that meaningful diversity embodies, major quality problems in the results are neglected, which, as we will show, can result in far-reaching forms of westernized cultural homogenization and epistemic injustice (Spivak, 1988).

For these reasons, in this paper, we do not echo the call for bridging the divide by simply digitizing and integrating all the world's languages into existing large-scale technological infrastructures. Instead, what we are concerned with, is the structural inequalities expressed in the disparity of linguistic representations, the causes and consequences of that inequality, and the question how it can be addressed in ways that do not end up contributing to neocolonial dynamics. The first of our paper's three contributions is thus to define and outline in more detail the phenomenon of language modeling bias. A resource or tool exhibits language modeling bias if, by design, it is not capable of adequately representing or processing certain languages while it is for others. As we show with several cases of lexical gaps, language modeling bias is closely related to a second key concept, linguistic diversity, which refers to linguistic constructs and ultimately ideas that are difficult to translate into certain languages. We argue that if we are to do justice to the more profound dimensions of diversity expressed in different languages and prevent epistemic injustice by means of language modeling

[1] According to https://meta.wikimedia.org/wiki/List_of_Wikipedias, retrieved on 1 May 2023.

bias, we need to pay attention to precisely these constructs and the cultural particularities they reveal.

In our second contribution, we detail how the causes of language modeling bias are in part a consequence of the flawed methods by which language technology is currently developed. In doing so, we refer primarily to the academic apparatus of knowledge and technology production, but also outline how this apparatus is intertwined with the private sector and its interests. To support our claim, we analyze how many databases and language processing systems that are purportedly multilingual have been developed from the perspective of a single language (English). Without a comprehensive understanding of linguistic diversity that goes beyond simply representing another language in a pre-existing model, they run the risk of only superficially filling a language gap, while on closer inspection being ineffective at supporting the values that closing the gap is intended to promote.

Our third contribution addresses the ethico-political implications of our analysis. We show how simplistic representations of diversity can lead to inevitably false representations of particular languages, which, when they penetrate previously under-served communities, can lead to dialectic dynamics by perpetuating existing or new forms of epistemic injustice, which we outline below in more detail. As an alternative, we make a case for a language resource development initiative we call *LiveLanguage,* which is grounded on rigorous co-design, thereby reflecting, supporting and accounting for diversity in a much more principled and systemic manner than any top-down approach can (Saad-Sulonen et al., 2018; Smith et al., 2021).

In line with these contributions, the paper is organized as follows. In Sect. "Diversity as an ethical norm", we define and discuss the notion of (linguistic) diversity as well as epistemic injustice, respectively. Section "Bias in language technology" is devoted to the definition of (language modeling) bias, the various forms it can take and its causes. Section "Ethical concerns with biases in language technology" tackles the normative consequences that follow from the language modeling bias we identified. Finally, in Sect. "Addressing epistemic injustice in language technology: the livelanguage initiative", we discuss approaches to co- and participatory design in language technology and clarify some of the conditions that we see need to be fulfilled in order to avoid ethical harms.

Diversity as an ethical norm

Although our point of departure is the normative one of protecting, promoting, and preserving diversity for the sake of epistemic justice, we are wary of the problems that come with naively celebrating it without proper conceptualization (Helm et al., 2022). Acknowledging that diversity is a moral-epistemic hybrid (Potthast, 2014), we differentiate between an understanding of it as a *descriptive* and a *normative* concept, to better distinguish between (a) the actual notions of difference that underlie our understanding of linguistic diversity as a design strategy, and (b) the values we associate with it as the objective of our work.

Diversity: a moral–epistemic hybrid

A closer look at the ethics policies of large tech companies reveals that while diversity is regularly listed as a core corporate value, it is often reduced to simplistic but easily measurable categories such as gender, race, or age. Ruha Benjamin aptly described this as "cosmetic diversity" (Benjamin, 2019, p. 24). Cosmetic diversity is problematic for several reasons. First, because it clouds our eyes to the ambiguity of diversity as an instrumental and thus conditional value. Second, because such portrayals often lead to a treatment of diversity as a resource that can be "exploited." Political philosopher Iris Young, however, warned already in the 1980 s against such capitalist appropriations of the concept, where diversity is instrumentalized as something that "enriches me" or as a means of optimally valorizing people. Instead, diversity is about how we can live together in an inclusive, participatory, and nondiscriminatory way (Young, 1990).

To clarify this difference, anthropologist Anna Tsing speaks of "meaningful diversity," that is, diversity that changes things, as opposed to scalable diversity, which accepts only what can be incorporated into pre-existing standards without further adaptation (Tsing, 2012). Tsing's distinction between meaningful and scalable diversity is instructive here, as it highlights exactly the difference that we want to point out when criticising current attempts to increase linguistic diversity in language technology, in ways that simply extend systems already in place. These attempts, we will show, fail to account for the more profound cultural and epistemological differences, which are incorporated within different languages and which, as we claim, should be at the heart of diversification efforts. This, however, requires much more profound adoptions all the way through the methodological, modeling, design, and implementation circle.

Linguistic diversity

As a design strategy, diversity helps define differences between entities, such as languages, and point out their unique features (e.g. words or expressions that cannot be translated easily into other languages, notions that only make sense to specific speaker communities). The terms *language diversity* and *linguistic diversity* are often used to refer to the

over seven thousand languages existing in the world, and to the wide-ranging differences among them (Giunchiglia et al., 2018). The association of diversity to language implies the preservation of the variedness of the world's linguistic landscape. In the field of linguistics, diversity is not a technical term and is therefore usually used in an informal way, with a few notable exceptions. Greenberg (1956) defined linguistic diversity as the probability of two persons speaking the same language in a certain geographic area. Rijkhoff et al. (1993), instead, apply the term (informally) to sets of languages, and understands the 'variedness' of the languages in terms of their genetic relationships.

With the aim of assessing instances of language technology in terms of their representation of linguistic diversity (or the lack thereof), we draw on the previous distinction between meaningful and scalable diversity and relate it to language technology. This helps to critically scrutinize existing attempts at closing the digital language divide: whether a given language technology does justice to the normative dimensions of diversity, representing the wide-ranging semantic and grammatical specificity's of the languages to which it is applied.

> A technology can be qualified as doing justice to linguistic diversity in a meaningful way if it is able to process and represent different linguistic means available in different languages even when the most well represented languages do not provide an equivalent means and thus can only indirectly or approximately express the idea.

The most straightforward examples of linguistic diversity are found in lexical semantics, in relation to the well-known phenomenon of untranslatability. One example from the domain of kinship (the diversity of which is well documented) is the Maori word *teina*: it means *elder brother* if it is pronounced by a male speaker, and *elder sister* if it is pronounced by a female. In translation to English, this concept can only be expressed in an approximate way. Another example is the phenomenon of *inalienable possession*, widely present in Native American and Australasian languages, where abstract—yet for us natural—concepts such as *mother* or *head* (as a body part) cannot be expressed as single words (free morphemes), but only together with their possessor (i.e. as the combination of two bound morphemes): *my mother*, *your head*.

Motivating our normative stance on the importance of properly dealing with diversity when building or expanding language technology, we claim that, for native speakers, such language-specific terms are often inextricably embedded in the local context. For a speaker in South India, choosing the correct term out of 16 possible terms to designate one's cousin—depending on gender, age, the mother's or father's side, etc.—is a basic requirement of politeness and culture, while in other languages, there is only one single term existing for cousin. Although kinship is a prime example of linguistic diversity, it can also be reflecting of geographical specifics of particular regions. For example, in the Italian Alps, the word *malga*, designating a typical mountain restaurant with no equivalent outside the Alpine region, is an important everyday term with a strong connection to south alpine tradition and culture.

From the perspective of computational linguists and engineers, in contrast, diversity represents a boundary beyond which algorithms do not scale. Given the persistent and increasing scaling pressures in the field, which we will outline in more detail in the next section, it is a well-understood temptation to simply ignore such long-tail phenomena and concentrate on the more high-level representation as an easy way to achieve scalable diversity. Yet, it is not always impossible to reconcile the engineer's inclination for automation with an accurate computational representation of linguistic diversity. One solution is to rely on the vast scientific data on linguistic typology produced by experts through the last century. Giunchiglia et al. (2017) used a quantified measure of the diversity of sets of languages for the prediction of the universality or specificity of linguistic phenomena. Khishigsuren et al. (2022) used results from in-depth, local field studies to better understand the meaning of family relations in order to produce accurate kinship terminologies in no less than 600 languages. In Bella et al. (2020), an about 10-thousand-word formal lexicon of Scottish Gaelic was co-created by local language experts, including locally specific terms not directly translatable to English or most other languages.

These examples show that the technical representation of meaningful linguistic diversity is not only a question of feasibility, but also one of normative orientation and the related priorizations leading to an intensified investment in engagements with local communities and co-creation efforts.

Epistemic (in)justice

We have already pointed out the importance of rigorous conceptual work for the meaning we attach to the normative concepts that guide our efforts. It is equally important to clarify what is lost or which kinds of harms are done when these norms are violated, that is, when diversity is simplified in such a way that its normative dimensions are eroded. In the introduction, we used the term "epistemic injustice" because it not only describes well the homogenization that can result from loss of diversity, but also situates that loss and the attached harms within a broader context of global inequalities.

The term "epistemic injustice" was introduced by philosopher Miranda Fricker (2009) and refers to a typology of injustice that is distinct from the injustice caused by the inequitable distribution of epistemic goods, such as educational

materials, books, or information technologies. It is therefore very useful in accurately understanding and naming the problems that arise when language modeling bias persists despite, or because of, the broad extension of language technologies to a variety of languages and communities. Rather than focusing on the issue of distribution of resources (Goldman, 2002), epistemic injustice, as understood here, addresses the harms that occur at a more subtle level when people are unequally valued in their capacity as bearers and practitioners of different forms of knowledge (Coady, 2010). According to Fricker's analysis, the most important forms of epistemic injustice include forms of exclusion and silencing, the systematic distortion or misrepresentation of certain people's meanings or contributions, and the undervaluing of their status in communicative practices.

Epistemic injustice also has a clear political connotation in that it disproportionately affects groups of people who are already disadvantaged because of their social identities, such as race, gender, class, or disability. In addition to the inequitable distribution of resources, epistemic injustice affects the ways in which knowledge and experiences are recognized, valued, or discredited by others. It manifests itself in two main forms: testimonial and hermeneutic injustice. The second of which is most relevant to the present case. Hermeneutic injustice refers to a situation in which a person or group is disadvantaged because their experiences or social realities are not acknowledged or understood due to a lack of concepts, vocabulary, or frameworks. In such cases, it may be difficult for individuals to articulate their experiences or seek redress because this particular, rather subtle but no less relevant form of injustice is not adequately recognized or understood by society (Fricker, 2009).

From an overarching perspective, the concept of epistemic injustice also needs to be situated historically, as it can be understood as a further development of Gayatri Chakravorty Spivak's notion of epistemic colonization (Spivak, 1988). Epistemic colonization refers to the processes by which one's culture's knowledge systems, beliefs, and ways of knowing are imposed on another culture or community, often as a direct, indirect, or late consequence of colonization or imperialism. This involves the domination of a particular theory of knowledge (in the present case, it may be a belief in the universal power of AI systems developed in the West) over others, often marginalizing or suppressing local knowledge systems and ways of understanding the world.

Epistemic injustice, understood as rooted in a history of epistemic colonization, can lead not only to individual but also to structural harm, as it is usually accompanied by a loss of cultural diversity and leads to a form of homogenization or violent cultural appropriation that ultimately benefits those who caused the injustice. In this way, existing power imbalances are perpetuated as imperialist knowledge becomes the standard against which all other knowledge is measured. This can entrench structural dependencies. Epistemic injustice, as we understand it here and use it to critically assess the effects caused by current initiatives to expand language technology, builds on historically established inequalities and need to be understood in this cronyism. In our view, counter-designs and strategies can only function if they take this broader context into account.

In what follows, we lay out how recent attempts to close the language gap through distributing epistemic resources but without accounting for meaningful diversity are at risk of contributing to epistemic injustice. To do so, we elaborate on what language modeling bias means as a counterpart to linguistic diversity, and why it is much more of an ethico-political matter than it might appear from a purely technological or linguistic perspective.

Bias in language technology

To understand how bias plays out in language technology, it is important to consider how linguistic bias, a well-researched area, gets interwoven with algorithmic bias, another well-researched area.

In the context of digital technology, the notion of *bias* has gained much attention and was prominently problematized as it has been identified as one of various sources of automated discrimination (Barocas and Selbst, 2016). So far, algorithmic bias has been used mainly to refer to patterns of stereotypes and preferences towards social groups, most often concerning learning-based language processing systems (Blodgett et al., 2020). In terms of social groups, studies have focused on gender, ethnicity, and race, but also other forms of bias (religion-related, age-related, political, etc.) (Friedman & Nissenbaum, 1996). To, then again, systematize linguistic bias, which is at the focus of communication studies, Hovy and Prabhumoye (2021) identify five sources: (1) the training data, (2) the annotation process, (3) the models, (4) the input representation process and (5) the research design for studying the biases. All these five sources we also find to be relevant when dealing with algorithmic bias in, say image recognition or ADM-systems (De-Arteaga et al., 2019a; Schwemmer et al., 2020).

As it can lead to various forms of harm, it is important to problematize both linguistic and algorithmic bias, and particularly their interplay. Yet, we also recognize that bias is omnipresent and that, even if it is usually associated with a negative connotation, it actually need not be harmful to diversity *per se*. For example, when affirmative action serves to counteract the unequal representation of otherwise marginalized groups, bias may well be intentional and desirable. Contrary to a blanket critique of bias as a phenomenon in itself, we accept that all knowledge, all insights, and even all data are situated, i.e. they always reflect a particular point

of view in space and time that is influenced by culture, history, politics, economics, epistemology, and so on (Haraway, 1988; Gitelman, 2013).

Unbiasedness is therefore a deceptive goal that, instead of solving social problems, reproduces problematic ideas, such as the unrealistic imaginary that technology can be neutral (Beer, 2017). It is therefore important to be upfront about when and for what reasons a certain bias is problematic and needs to be combated, and that this combating does usually not lead to no bias, but to a different, ideally more just bias (Harding, 1995). Linguistic bias, for example, is harmful to diversity *when* it perpetuates existing or produces new forms of hermeneutic injustices related to already vulnerable, and/or marginalized language communities. Such bias calls for counteraction. When such linguistic bias is then reproduced through LLMs that are geared toward the correct representation of languages of colonial powers, but disregard the particularities of other languages that are also spoken by many people or are at risk of extinction, this demands change. To enact such change sustainably, it is instrumental to invest work into unraveling how linguistic bias and algorithmic bias interact to emerge as a new subform of bias, which we call language modeling bias.

Language modeling bias

The subject of language modeling bias are not just languages *per se* but also the design of language technology: corpora, lexical databases, dictionaries, machine translation systems, word vector models, etc. Language modeling bias is present in all of them, but it is easiest to observe with respect to multilingual resources and tools, where the relative correctness and completeness for each language can be observed and compared. We define it as follows:

> Language modeling bias is observed when the technology, by design, represents, interprets, or processes content less accurately in certain languages than in others, thereby forcing speakers of the disadvantaged language to simplify or adapt their communication, (self-)representation, and expression when using that technology to fit the default incorporated in the privileged language.

Bias manifests itself through linguistic or cultural inaccuracies in the way a language is processed or represented. By emphasising the *by-design* aspect of language modeling bias, our definition is deliberately focusing on the representational, rather than the allocational harm of bias in language technology. Thus, language modeling bias is not only concerned with the scarcity of data on certain languages or with biases within linguistic devices, but rather with how these biases are amplified through structural bias built into language processing algorithms, representational models, resources, or methodologies.

We thus situate language modeling bias as a specific form of *algorithmic bias* that is observed in language technology. We differentiate it, from other amply studied forms of algorithmic bias in language-based AI, such as semantic representation bias, that are not primarily linguistically defined De-Arteaga et al. (2019b). In opposition, the direct subjects of language modeling bias are in fact languages, dialects, and sociolects, while its indirect subjects are, of course, the speakers themselves. A second, crucial distinguishing feature from other forms of bias is that language modeling bias concerns *technology design*: inherent limitations within the structure of language databases, neural AI systems, and language processing algorithms. We clearly distinguish this issue from out-of-scope problems related to the underlying data (corpora), frequently included under the umbrella term of *algorithmic bias*: data availability, such as differences in the sizes of training corpora between well-resourced and under-resourced languages, or data quality, such as sociocultural stereotypes encoded within training corpora. Focusing on these problems would lead to solutions that confirm to scalabe as opposed to meaningful ideas of diversity, such as simply generating more language data to be fed into predefined LLMs.

The social groups affected by language modeling bias are clearly the communities of speakers of underrepresented languages, however heterogeneous they may be otherwise (according to social status, culture, gender, race, ethnicity, religion, etc.). Being the native or second-language speaker of a language variety determines one's access to information, and the language technology that enables this access affects one's ability to communicate, on the Web or elsewhere. To our knowledge, the term *language modeling bias* has not been used as a analytical device or design strategy in any way similar to ours while many of the underlying neocolonial mechanisms have, however, been pointed out (Bird, 2022; Schwartz, 2022).

In terms of actual bias in AI systems and data, research concerning the representation (or the lack thereof) of the vernaculars of social groups within language resources is closest to ours. Here, however, we want to go a step further in pointing out how, both in the field of engineering and technology advancement as well as in ethics, policy and development aid, the language communities themselves are left out of the process. These attempts or projects, with Aradau and Blanke (2022) can be described as techniques of governing emerging technology, which while striving for diversity as one of their goals, turn those most affected by the results into what philosopher Jacques Rancière has called the "part of those who have no part" (Ranciere, 1998, p. 30). It is the technology developers and designers residing in the big companies as well as influential academic institutions

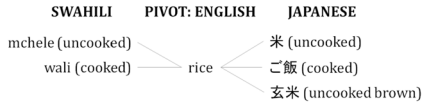

Fig. 1 Biased cross-lingual mapping of words about various forms of 'rice' from a popular multilingual lexical database

that fashion themselves as the experts who are called upon to embed linguistic diversity into their tools and expand them under the normative guise of inclusion. However, in the process, language modeling bias is reproduced because the Western perspective is taken as the norm and the subjects of diversity remain at the outside. It is this form of exclusion which at the same time allows for and is mobilized by investments in scalable diversity (Fig. 1).

Forms of language modeling bias

In the following we provide examples of language modeling bias from mainstream AI-based language technology: neural language models, machine translation systems, and multilingual lexical databases.

Neural language models

The general trend of AI-based (neural) language technology is to rely on as little prior knowledge about languages as possible, instead obtaining all such knowledge through corpus-based learning. While such a design avoids any obvious algorithmic bias towards any particular language, it creates a strong dependency on the quality and the size of the training corpus. The well-known consequence of this is a preference of mainstream AI towards the languages of dominant cultures with a strong web presence, with gigabyte-sized pre-trained *large language models* that have led to not one but a long series of breakthrough improvements on language understanding and generation tasks. The same technology, due to the lack of corpora, provides low-quality results, or none at all, to under-resourced languages.

While the dependence of neural language models on large training corpora appears to be a transparent and seemingly language-agnostic constraint, the architectural choices underlying such models can still lead to language modeling bias. White and Cotterell (2021) show that the word prediction performance of the *Long Short-Term Memory* (LSTM) neural network architecture is less sensitive to word order than that of the Transformer architecture. As they experimentally show, the Transformer appears to have a bias towards the (rarely occurring) verb–subject–object (VSO) word order, while showing lower performance on the (very frequent) SOV and SVO word orders, all other parameters being equal. Moving to morphology, Zevallos and Bel (2023) study subword-frequency-based, language-agnostic tokenization (i.e. word splitting) algorithms, such as *Byte Pair Encoding* (Sennrich et al., 2015), that are typically used to preprocess corpora fed to large language models. They experimentally show that such methods tend to train slower on morphologically complex (synthetic, agglutinate) languages, meaning that more training data are required for these languages to achieve the same performance on downstream language understanding tasks. Replacing the language-agnostic tokenization algorithm by language-specific morphological segmentation allows language models to train more efficiently over smaller corpora.

Machine translation

Machine translation (MT) is the flagship task of AI-based language technology. Without claiming to be exhaustive, we point out three aspects of current MT technologies where language modeling bias can be observed: the non-handling of untranslatability, the variedness of vocabulary and grammar, and the use of a pivot language.

Today's top MT systems, such as DeepL and Google Translate, make systematic mistakes over untranslatable terms, betraying the fact that this phenomenon is not specifically addressed by these tools. The screenshots (a) and (b) in Fig. 2, taken from a mainstream machine translator, show examples of erroneous translations due to untranslatability. As reported by (Khishigsuren et al., 2022), when translating the English sentence *My brother is three years younger than me* to Hungarian, Mongolian, Korean, or Japanese, syntactically correct yet semantically absurd results are obtained:

Hungarian: **A bátyám három évvel fiatalabb nálam.*
Japanese: *私の兄は私より3歳年下です.
Korean: *형은 나보다 세 살 아래다.
Mongolian: *Ах маань надаас гурван насаар дүү.

These languages either have no equivalent word for *brother* (as in Mongolian) or, when they do, the equivalent word is rare (as *fiútestvér* in Hungarian). Based on training corpus frequencies, the MT system ends up choosing a semantically unsuitable word, such as *bátyám* meaning *my elder brother*, resulting in *My elder brother is three years younger than me.*

A second form of bias concerns the variedness of vocabulary and grammar in MT output. Vanmassenhove et al. (2021) quantitatively compare the lexical and grammatical 'richness' of original and machine-translated text. They report that both lexicon and morphology tend to become poorer in machine-translated text with respect to the original (untranslated) corpora: for example, features of number or gender for nouns tend to decrease. This is a form of language modeling bias against morphologically rich languages.

A third form of language modeling bias in MT is their use of English as a pivot language when translating between

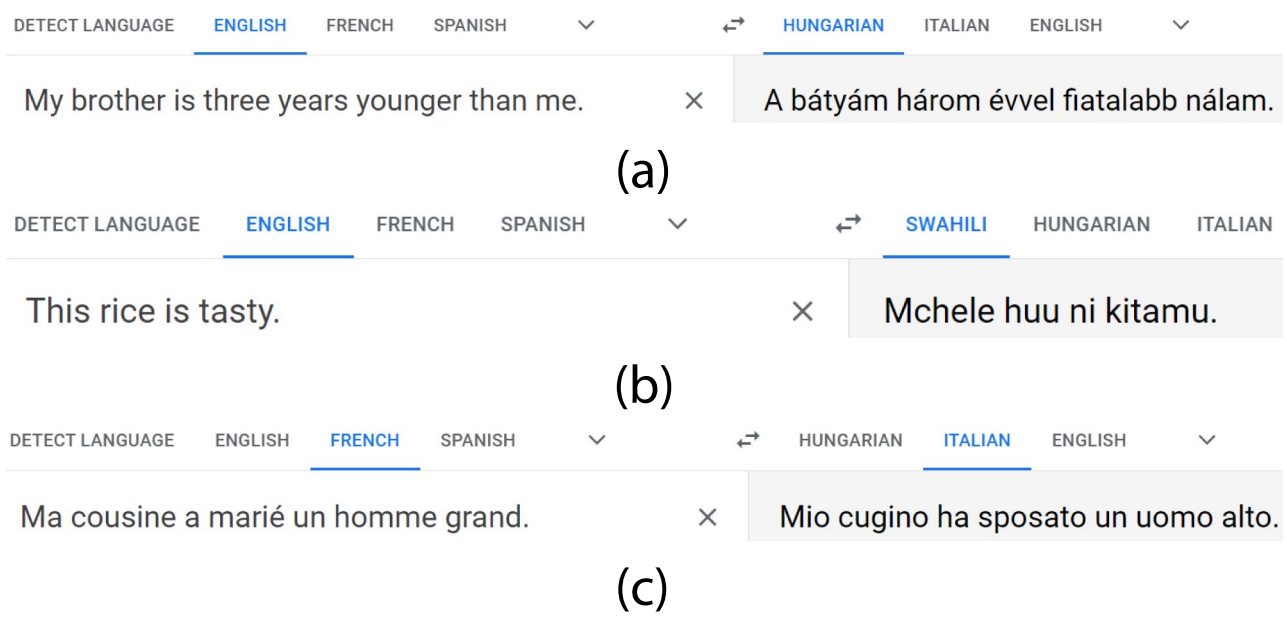

Fig. 2 Examples of language modeling bias in machine translation. **a** The lack of an equivalent common word in Hungarian for *brother* results in an erroneous translation meaning *my elder brother is three years younger than me*. **b** The lack of an equivalent term for rice in Swahili results in an erroneous translation meaning *this raw rice is tasty*. **c** The systematic use of English as pivot language results in an erroneous change of gender when translating between French and Italian

two languages other than English. This practice is explained by the relative scarcity of bilingual training corpora for such language pairs, as well as scalability. Example (c) in Fig. 2 shows the case of French-to-Italian translation of a sentence meaning *my (female) cousin married a tall man*. While French and Italian (as do most languages) use different words for male and female cousins (*cousin/cousine*, *cugino/cugina*), English does not. The result is that the gender of the cousin is 'lost in translation' and, as a form of combined linguistic and gender bias, it appears as a male in the translation.

Multilingual lexical databases

As a generalisation of bilingual dictionaries, the 2000s saw the appearance of *multilingual lexical databases* that map words, based on their meanings, across a large number of languages. While these resources proved to be extremely useful in cross-lingual applications, looking under the hood—into their underlying models of lexical meaning—reveals varying levels of limited expressivity and bias.

As shown by Giunchiglia et al. (2023), some of these multilingual databases interconnect words from hundreds of languages, mapping the words of each language to a collection of roughly 100 thousand word meanings (so-called *synsets*) obtained from the English Princeton WordNet (Miller, 1998). On the one hand, this choice makes practical sense, as among all similar resources, WordNet provides by far the broadest and most precise semantic coverage. On the other hand, using the lexical concepts of WordNet to describe the lexicons of all other languages results in a strong bias towards the English language and Anglo-Saxon culture in general, as the expressivity of the database is limited to notions for which a word exists in English (Giunchiglia et al., 2023; Bella et al., 2022). Figure 1 provides a simple example from the food domain, known to be culturally, and thus also linguistically, diverse. It shows how a biased lexical database maps together words in Swahili and Japanese meaning *uncooked rice*, *cooked rice*, and *uncooked brown rice*. The degree of information loss is flagrant: while both Swahili and Japanese provide fine-grained lexicalizations about the various forms of rice, the many-to-many mapping that results from passing through English masks all fine-grained differences, resulting in both a loss of detail and incorrect translations when one moves from Swahili to Japanese or vice versa. The diversity-diminishing bias is also found in other domains that are well-known to be diverse across languages: family relationships, school systems, etc.

Methodological causes of language modeling bias

Because bias is most problematic when it perpetuates existing power relations that can contribute to far reaching harms such as hermeneutic injustice, any critique of technological bias should ideally include at least a brief genealogy of the origins of bias and the ways in which different social groups are harmed or benefited in different ways. The following subsection highlights how well-intended but unreflective

attitudes in computational linguistics contribute to the creation of language technologies that are adverse to meaningful diversity. It also reveals that computational linguistics has not developed in isolation, but is situated within epistemic hierarchies that are both reflective of and contributing to socio-economic power asymmetries. Analyzing this kind of situatedness of research cultures is crucial for clarifying why focusing on language inclusion and expansion is not enough to promote diversity, but may, when being scaled, ultimately even reproduce entrenched dynamics of epistemic injustice.

In the last 50 years, research in Computer Science has been dominated by a strong Anglo-Saxon influence, reflecting turn-of-century power dynamics. In Computational Linguistics, this bias was apparent across all prestigious publications, conferences, and journals of the field: an unspoken convention required for research to be considered as competent to be applied and demonstrated in English. Thus, English has not only been the *lingua franca* of scientific communication, but also the *de facto* standard subject matter of research. This is not to say that scientific results on other languages were not published, but they were generally considered by the community (paper reviewers, journal editors, etc.) as 'language-specific' and therefore less likely to be relevant to a wide audience. Publications about languages other than English were relegated to second-tier or niche journals and venues. Schwartz (2022) reports that between 2013 and 2021, 83% of papers accepted at ACL—the flagship conference in Computational Linguistics—were explicitly or implicitly about English and 97% were about Indo-European languages.

Despite these numbers, the 2010s saw an emerging interest in multilingual language technology, and of a new research sub-field targeting 'low-resource' (or 'under-resourced') languages, previously neglected by mainstream research. This change of scope is tightly related to the blazing progress of deep-learning-based AI on English (and also on some other well-supported languages such as Spanish or Chinese). Problems that were earlier considered as exceedingly hard, such as machine translation, have suddenly been solved with impressive results. For researchers, the 'major' languages were not providing suitably interesting challenges anymore, apart from incremental research pushing the accuracy boundary. Low-resource languages seemed like a promising horizon.

The new fascination with 'low-resource languages' does not mean that, say, Mongolian speech synthesis has suddenly become of mainstream scientific interest. In line with the 'zero-shot' data-driven ethos (Bird, 2020) of recent deep AI research that shuns any use of prior results from linguists and field workers, low-resource language research is only worthy of a top publication as long as (1) it provides a solution for multiple, preferably tens or hundreds of languages at the same time; (2) it involves mainstream AI technology, i.e. neural networks; and (3) it requires very little to no knowledge from experts or speakers of the languages targeted. The typical low-resource research contribution thus scrapes web content, such as Wikipedia pages, written in the languages in question, often without any understanding of their quality or content (Lignos et al., 2022). It then trains or fine-tunes deep learning models based on the data, and finally demonstrates a few percentages of increase in quality (precision, recall, BLEU, etc.) over one of the standard tasks in computational linguistics, such as named entity recognition or machine translation, against corpora that the researchers themselves cannot read. This practice is certainly not in line with what we earlier described as accounting for meaningful diversity.

Simultaneously, many highly populated but under-resourced Global South countries were identified by high-tech companies (and digital platforms in particular) as still unsatisfied markets with a potential for data scraping and infrastructural advancements. What happened during the following 10 years has been described as a 'race' in which digital platforms swamped African and South East Asian countries, in order to be the first to secure the loyalty of wast new customer bases (Arora, 2019; Benjamin, 2019).

Also, in the field of technology ethics, Silicon Valley has set agendas over the past decade by pumping large amounts of money into an academic system that otherwise faces scarcity measures and budget cuts (Ochigame, 2019). This has led to two types of ethics increasingly taking hold: firstly, the type that embraces the notion that it is primarily more technology, and in particular the expansion of AI, whereby existing problems can be solved, and secondly, the type of ethics that can be easily transferred into existing systems and infrastructures and from there automated and implemented *en masse*. Both can be demonstrated in the often overly simplistic ways in which racial and gender biases are tackled by developing fairness measures, and can equally be mirrored for the appropriation and handling of calls for better acknowledgment of diversity.

The industrial appropriation of academic ethics research and its influence on the respective notions of justice as fairness and diversity as demographics go hand in hand with the broader ranging imaginary that large scale technological innovation will serve as a panacea for wide-ranging problems (Pfotenhauer and Jasanoff, 2017). Following Anna Tsing, however, we understand scalability not as an intrinsic property of a solution or product (of any kind), but as something that stems from emphasizing certain aspects at the cost of others. Ironically, then, for an innovation to be scalable, it must be designed to reduce the complexity of a problem and its associated solution to isolated parameters that can fairly easily be abstracted from the context of the specific domain or community for which it was developed (Engel, 2016). This abstraction work makes the innovation generalizable

and thus scalable in that the number of languages can be significantly increased without major adaptation (Tsing, 2012). This is exactly what is happening when existing neural language technology is applied indiscriminately to any language without adaptation.

These problems are exacerbated against the backdrop of a postcolonial computer culture (Irani et al., 2010). In this context, recent Data4D efforts have been criticized, not only in terms of their "white savior" ethos, but in some cases even to the point of using development goals as free riders to invade vulnerable populations and extract their data (Taylor and Broeders, 2015). Whatever the intentions, the choice of problems to focus on is often driven by either the incentives of academic communities or by industry pressure or, most likely, a combination thereof. This leads to a gap between the solutions offered and the diverse needs of communities.

Ethical concerns with biases in language technology

The consequences of research being done under these conditions raise a multitude of ethical concerns with regard to potential epistemic injustice being done. Most of these concerns are related to the rather ill-defined attempts to promote diversity by adopting the top-down scalable solutions that AI-approaches warrant and which, by the nature of their design, can only respond to a simplistic idea of diversity. With Western researchers unilaterally setting developmental goals and providing technological solutions to reach them, they effectively and most ironically, silence the actual speakers. This silencing does not regard the lexical representation or distribution of epistemic goods, which may in fact be increased, but the types of problem definitions and the corresponding designs of technical solutions.

Yet, language resources that are, at least partly, handcrafted and co-designed, are rarely deemed competitive because they are much harder to scale as they are by definition not generalizable. If technical innovation is, however, narrowly defined as the expansion of AI via Neural Language Models, and social innovation as the scaling of cosmetic diversity, then this will not only lead to a neglect of small languages. It will also affect large language communities when the social realities and worldviews anchored in their language cannot be expressed through dominant anglo-centric models. Given the importance and possibilities of language technology in the struggle for hermeneutic justice, defined as the equal recognition of distinct socio-culturally situated experiences or realities, there is a strong case to be made for the socio-technical innovation potential of co-designed and customized systems that can do justice to linguistic diversity.

Despite this potential, critical commentators on AI language technology point out how well-intended research goals such as "technology-based revitalization" regularly misinterpret the needs of local communities (Bender et al., 2021; Bird, 2022). In most cases, native speakers are not involved in the process, or if they are, they are taking on subordinate roles such as commentator, validator, tester, or worse, data extractor (Helm et al., 2023). Instead of co-creating on an equal footing, in many cases the analytical, high-level work is done in technology labs of Western universities or companies, where the languages being studied are often not even understood by the people working on them, let alone the cultures they represent (Arora, 2016). Sometimes they do not even know if they are using the right language, as observed in the case of automated Wikipedia scraping (Lignos et al., 2022).

Added to this is the neglect of the meaning and relevance of language variations, which goes beyond the mere transmission of information (Bender et al., 2021). The Kenyan writer and scholar Ngugi wa Thiong'o (1986), for example, has done extensive research on the cultural diversity embedded in Kenya's multilingual heritage. Language conveys the situatedness of knowledge. It influences how we see our world, which visions we follow, how we perceive colors, tastes, time. In Kiswahili, Arabic and Amharic, for example, time is measured not only from one hour to the next, but in relation to sunrise and sunset. Thus, 7 a.m. depends on the season because it is always the first hour after sunrise. This understanding, which better adapts social life to the dynamic rhythm of the year, does not translate one-to-one into the much less dynamic but more definite Anglo-American system for communicating about time, and it would be lost if we all had to squeeze ourselves into English. It is this kind of situated knowledge embedded within languages, which cannot be captured by one-size-fits-all design, nor by AI, no matter how technically sophisticated. It is something that has to be done by people who know what they are talking about, literally. These dimensions are in danger of getting lost to a culture that limits its normative horizon to large scale technical innovation and expansion as ends in themselves.

Given these differences, which go beyond lexical representation, we contend that it is unfair to require all of us to conform to communication norms that have emerged from the perspective of English-speaking (or, increasingly, Chinese-speaking) users if we are to take advantage of the opportunities that language technologies provide. In other words: can we accept language modeling biases as a fact of the digital age, or do we consider them ethically unacceptable? Following our reasoning, the latter is obviously the case. Having analyzed the various forms language modeling can take, we argue that just as entrenched gender biases can be found in image recognition systems, with far-reaching consequences for equal opportunities for non-cis people,

this is the case for linguistic biases in language technology. Except that the latter form has to do with the misrepresentation or disregard of cultural or "social factors" embodied in and expressed through language (Hovy & Yang, 2021). This disregard does not necessarily manifest itself in direct acts of discrimination, but in systemic forms of hermeneutic injustice. While these forms are difficult to grasp in their wider implications, in the most severe instances, they do reveal themselves in their harmful impacts on the life chances of individuals (as we can already see from the dramatic consequences that errors in automatic translation can have in the context of asylum procedures, (Bhuiyan, 2023)).

The challenge now is to understand these errors early on not just as incidental glitches, but as the result of broader, socio-historical inequalities that manifest themselves in the performance of widespread language technologies (Broussard, 2023). When lexical gaps caused by biased translation technologies become a normal condition of multilingual communication, then this will lead to a situation where one person can express herself perfectly well in her own language, while another one is limited in her expression of her social realities and the experiences that emerge from these. This is nothing else than hermeneutic injustice playing out in concrete practice. To defend this rather strong claim, it is important to consider the power relations rooted in a colonial history marked not only by necropolitical but also empistemic violence. The political theorist Ali Mazrui and Mazrui (1999) has studied the influence of English on African culture and argues that this influence changes the self-perceptions and social practices of African peoples. This need not be bad *per se*. Creole creates something new that we can greatly appreciate; cultural encounters can be enriching and broaden perspectives. The problem, then, is not cultural mixing and matching as such; on the contrary. Rather, it is the dominance of certain cultures over the others, that tends to be reflected or even perpetuated in AI language technology and which needs to be overcome if we are to promote epistemic justice.

Addressing epistemic injustice in language technology: the Live Language initiative

A case presenting a specific effort for embedding values in language technology is the *LiveLanguage* initiative (Bella et al., 2022). LiveLanguage combines diversity-aware design (Helm et al., 2022) with a collaborative resource development methodology that strives to promote epistemic justice for marginalized language communities.

In terms of technology, LiveLanguage focuses on the development of diversity-aware lexico-semantic resources. The *Universal Knowledge Core* (UKC) lexical database[2] covers the lexicons of over 2,000 languages (to varying degrees of completeness). What makes the UKC aware of meaningful diversity is its simultaneous representation of, on the one hand, what is shared across languages and cultures, such as words with equivalent or similar meanings or a common etymology and, on the other hand, what makes them different: untranslatable terms, lexical gaps, or language-specific grammar (see Fig. 3 for an example: languages that do and that do not lexicalize the family relationship of *female sibling*). It is through the integrated representation of cross-linguistic unity and diversity, both on the surface and in semantics, that the UKC confronts language modeling bias within lexical resources (Giunchiglia et al., 2023).

The diversity-aware lexicons, contained in the UKC in the form of an extensive lexico-semantic knowledge graph, are freely downloadable from the *LiveLanguage data catalog*.[3] They can be useful as reference knowledge for the evaluation of AI applications, and as input resources in order to complement corpus-based training. The use of lexical databases as reference knowledge has a long tradition in computational linguistics, such as in word sense disambiguation where lexicons are used as catalogues of word senses (Agirre & Edmonds, 2007). More specifically, LiveLanguage data was used to perform meaning-level evaluation of machine translation systems over hard-to-translate sentences (Khishigsuren et al., 2022). Likewise, cross-lingual cognate pairs from the UKC (Batsuren, Bella and Giunchiglia, 2022) have been employed to evaluate word embedding models (Zouhar et al., 2023).

LiveLanguage and the UKC integrate both existing third-party resources and linguistic data collected through collaborations with universities. Examples of such collaborations include Mongolia (Batsuren et al., 2019), Scotland (Bella et al., 2020), India (Chandran Nair et al., 2022), Palestine (Khalilia et al., 2023), and South Africa (Dibitso et al., 2019). Striving to ensure that such collaborations are beneficial to local speaker communities and to avoid exploitative practices, LiveLanguage collaborations adopt a methodology based on co-creation and local empowerment, with the following characteristics: (a) representatives of local communities are leading the formulation of problems and needs, as well as the subsequent specifications of the language resources to be developed; (b) tools, infrastructure, and know-how are provided to local communities if needed, in order to embed solutions sustainably; (c) intellectual property rights stay with the local community; (d) language resources are integrated into the global LiveLanguage

[2] http://ukc.datascientia.eu.

[3] https://datascientiafoundation.github.io/LiveLanguage/.

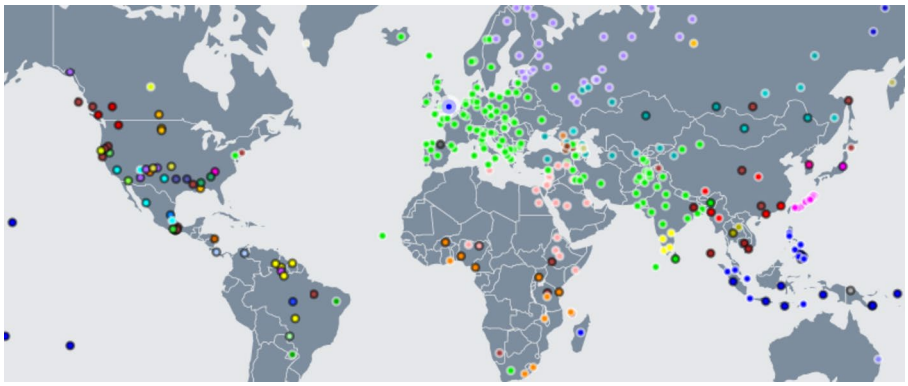

Fig. 3 Screenshot from the website of the UKC lexical database, showing languages that lexicalize the concept of *female sibling* (white-contoured dots) and those where it is known to be a lexical gap (black-contoured dots)

ecosystem, giving worldwide visibility to the results through the UKC database and the LiveLanguage data catalog.

That said, it needs to be noted that accounting for diversity and power asymmetries in language technology is not a fixed state, but a process and situated procedure which requires continual adaption to the variety of linguistic phenomena and different communities' needs. Therefore, we embrace the contributions of design anthropologists such as Smith et al. (2021) who advocate mutual learning and thus consider all participants in the process simultaneously as researchers and beneficiaries. Specifically with regard to language technology, we echo Bird's call for a focus on knowledge transfer beyond language, as generational loss of knowledge about local history, practices, etc. is often one of the main reasons for interest in language preservation, and gives rise to deliberate promotion of digitization.

Conclusions

In this paper, we have shown that simply applying existing language technology to ever larger sets of languages does not automatically serve the goal of bridging the digital language divide, as technology does not always generalize across languages. We argue that current technological approaches to addressing the "low-resource language problem" can even be detrimental rather than beneficial from the point of view of preserving linguistic diversity. Profound differences lie not only within diverse grammatical structures but also across people's social practices, worldviews, and situated knowledges embedded within linguistic expression. Through the LiveLanguage initiative, we try to make such differences manifest by formally representing cross-lingual diversity in both grammar and semantics, for domains such as kinship relations, educational systems, color, time, or food.

Furthermore, we point out that problems of ethnocentric language technology development are rooted within a colonial past, which is still potent today. Given these circumstances, there is an urgent need to be aware not only of the well-known forms of linguistic bias in AI systems that reproduce human biases encoded in large web corpora, but also to pay due attention to the language modeling bias that stems from language technology design itself. As we show, technological expansion that is based on biased tools is not only detrimental in terms of inadequate description of language, but actually contributes to a form of injustice that we identify as epistemic, or, more precisely, hermeneutic in its form and effect. This form of injustice is not aimed at the distribution of epistemic goods, which is indeed encouraged by recent efforts to expand multilingual language technologies. Rather, it is about the lack of recognition of certain forms of knowledge, modes of expression, and social realities that are evident in the diversity-related phenomena we have identified. This form of injustice is not only problematic in itself, but also troubling in that it can be understood as an extension of colonial domination.

In light of these criticisms, we conclude that any efforts to extend existing language technologies that are not based on a rigorous approach to co-creation with the language communities in question should be fundamentally re-framed. By rigorous, we attribute approaches based on a critical stance toward the privileges of whiteness that avoids any kind of white *savoir-faire* and instead conceives of the process as an opportunity for mutual learning in which neither party is superior to the other. It is this attitude and its related practices that will contribute to promoting and maintaining meaningful, as opposed to cosmetic diversity.

References

Agirre, E., & Edmonds, P. (2007). *Word sense disambiguation: Algorithms and applications*. Springer.

Aradau, C., & Blanke, T. (2022). *Algorithmic reason: The new government of self and other*. Oxford University Press.

Arora, P. (2016). Bottom of the data pyramid: Big data and the global south. *International Journal of Communication, 10*(1), 1–19.

Arora, P. (2019). *The next billion users: Digital life beyond the west*. Harvard University Press.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(3), 671–732.

Batsuren, K., Ganbold, A., Chagnaa, A., Giunchiglia, F. (2019). Building the mongolian wordnet. In: Proceedings of the 10th Global Wordnet Conference (pp.238–244).

Batsuren, K., Bella, G., & Giunchiglia, F. (2022). A large and evolving cognate database. *Language Resources and Evaluation, 56*(1), 165–189.

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society, 20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Bella, G., Batsuren, K., Khishigsuren, T., Giunchiglia, F. (2022). Linguistic diversity and bias in online dictionaries. University of Bayreuth African Studies Online,173.

Bella, G., Byambadorj, E., Chandrashekar, Y., Batsuren, K., Cheema, D., Giunchiglia, F. (2022). Language diversity: Visible to humans, exploitableby machines. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (pp. 156–165).

Bella, G., McNeill, F., Gorman, R., Donnaíle, C.Ó., MacDonald, K., Chandrashekar, Y., Giunchiglia, F. (2020). A major wordnet for a minority language: Scottish gaelic. In: Proceedings of the 12th Language Resources and Evaluation Conference (pp. 2812–2818).

Bender, E. M., Gebru, T., McMillan-Major, A., Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Proceedings of the 2021 acm conference on fairness, accountability, and transparency (p. 610–623). New York, NY, USA: Association for Computing Machinery. Retrieved from https://dl.acm.org/doi/10.1145/3442188.3445922 10.1145/3442188.3445922

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code (1. edition ed.). Polity.

Bhuiyan, J. (2023, September). Lost in ai translation: growing reliance on language apps jeopardizes some asylum applications. The Guardian. Retrieved from https://www.theguardian.com/us-news/2023/sep/07/asylumseekers-ai-translation-apps

Bird, S. (2020, December). Decolonising speech and language technology. Proceedings of the 28th international conference on computational linguistics (pp. 3504–3519). Barcelona, Spain (Online): International Committee on Computational Linguistics. Retrieved from https://aclanthology.org/2020.colingmain.313 10.18653/v1/2020.coling-main.313

Bird, S. (2022, May). Local languages, third spaces, and other high-resource scenarios. Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: Long papers) (pp. 7817–7829). Dublin, Ireland: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.acl-long.539 10.18653/v1/2022.acl-long.539

Blodgett, S.L., Barocas, S., Daumé III, H., Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in nlp. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5454–5476).

Broussard, M. (2023). *More than a glitch: Confronting race, gender, and ability bias in tech*. The MIT Press.

Chandran Nair, N., Velayuthan, R.S., Chandrashekar, Y., Bella, G., Giunchiglia, F. (2022, June). IndoUKC: A concept-centered Indian multilingual lexicalresource. Proceedings of the Thirteenth Language Resources and Evaluation Conference (pp. 2833–2840). Marseille, France: European Language Resources Association. Retrieved from https://aclanthology.org/2022.lrec-1.303

Coady, D. (2010). Two concepts of epistemic injustice. *Episteme, 7*(2), 101–113. https://doi.org/10.3366/epi.2010.0001

De-Arteaga, M., Romanov, A., Wallach, H., Chayes, J., Borgs, C., Chouldechova, A., . . . Kalai, A.T. (2019a). Bias in bios: A case study of semantic representation bias in a high-stakes setting. , 120–128. Retrieved from https://doi.org/10.1145/3287560.3287572

De-Arteaga, M., Romanov, A., Wallach, H., Chayes, J., Borgs, C., Chouldechova, A., . . . Kalai, A.T. (2019b). Bias in bios: A case study of semantic representation bias in a high-stakes setting. Proceedings of the Conference on Fairness, Accountability, and Transparency (p. 120–128). Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3287560.3287572 10.1145/3287560.3287572

Dibitso, M. A., Owolawi, P. A., Ojo, S. O. (2019). Context-driven corpus-based model for automatic text segmentation and part of speech tagging in setswana using opennlp tool. Modeling and using context: 11th International and Interdisciplinary Conference, Context 2019, November 20–22, 2019, proceedings 11 (pp. 62–73).

Engel, J. S. (2016). *Global clusters of innovation: Entrepreneurial engines of economic growth around the world (Reprint* (edition). Edward Elgar Pub.

Fricker, M. (2009). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems, 14*(3), 330–347. https://doi.org/10.1145/230538.230561

Gitelman, L. (2013). *Raw data is an oxymoron*. MIT Press.

Giunchiglia, F., Batsuren, K., Bella, G. (2017). Understanding and exploiting language diversity. Ijcai (pp. 4009–4017).

Giunchiglia, F., Batsuren, K., Freihat, A. A. (2018). One world–seven thousand languages. Proceedings 19th International Conference on Computational Linguistics and Intelligent Text Processing, Cicling2018, (pp. 18-24) March 2018.

Giunchiglia, F., Bella, G., Nair, N. C., Chi, Y., & Xu, H. (2023). Representing interlingual meaning in lexical databases. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-023-10427-1

Goldman, A. I. (2002). *51the unity of the epistemic virtues. Pathways to knowledge: Private and Ublic*. In Pathways to knowledge: Oxford University Press.

Greenberg, J. H. (1956). The measurement of linguistic diversity. *Language, 32*(1), 109–115.

Haraway, D. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies, 14*(3), 575. https://doi.org/10.2307/3178066

Harding, S. (1995). Strong objectivity: A response to the new objectivity question. *Synthese, 104*(3), 331–349.

Helm, P., Michael, L., Schelenz, L. (2022, Jul). Diversity by design? balancing the inclusion and protection of users in an online social platform. Proceedings of the 2022 aaai/acm Conference on ai, Ethics, and Society (p. 324–334). Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3514094.3534149 10.1145/3514094.3534149

Helm, P., de Götzen, A., Cernuzzi, L., Hume, A., Diwakar, S., Ruiz Correa, S., & Gatica-Perez, D. (2023). Diversity and neocolonialism in big data research: Avoiding extractivism while struggling with paternalism. *Big Data & Society*. https://doi.org/10.1177/20539517231206802

Hovy, D., & Yang, D. (2021, June). The importance of modeling social factors of language: Theory and practice. K. Toutanova et al. (Eds.), Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human language technologies (pp. 588–602). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.naacl-main.49 10.18653/v1/2021.naacl-main.49

Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass, 15*(8), e12432. https://doi.org/10.1111/lnc3.12432

Irani, L., Vertesi, J., Dourish, P., Philip, K., Grinter, R.E. (2010, Apr). Postcolonial computing: a lens on design and development. Proceedings of the Sigchi Conference on Human Factors in Computing Systems (p. 1311–1320). Association for Computing Machinery. Retrieved from https://doi.org/10.1145/1753326.175352210.1145/1753326.1753522

Joshi, P., Santy, S., Budhiraja, A., Bali, K., Choudhury, M. (2020, July). The state and fate of linguistic diversity and inclusion in the NLP world. D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 6282–6293). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2020.acl-main.560 10.18653/v1/2020.acl-main.560

Khalilia, H., Bella, G., Freihat, A.A., Darma, S., Giunchiglia, F. (2023). Lexical diversity in kinship across languages and dialects. To appear in Frontiers in Psychology, special issue on the adaptive value of language diversity. https://arxiv.org/abs/2308.13056 [cs. CL]

Khishigsuren, T., Bella, G., Batsuren, K., Freihat, A.A., Nair, N.C., Ganbold, A., Giunchiglia, F. (2022). Using linguistic typology to enrich multilingual lexicons: the case of lexical gaps in kinship. arXiv preprint arXiv:2204.05049.

Kornai, A. (2013). Digital language death. *PloS one, 8*(10), e77056.

Lignos, C., Holley, N., Palen-Michel, C., Sälevä, J. (2022, May). Toward more meaningful resources for lower-resourced languages. Findings of the association for computational linguistics: Acl 2022 (pp. 523–532). Dublin, Ireland: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.findings-acl.44 10.18653/v1/2022.findings-acl.44

Mazrui, A. M., & Mazrui, A. A. (1999). *The political culture of language: Swahili, society and the state*. Global Academic Publishing.

Miller, G. A. (1998). *Wordnet: An electronic lexical database*. MIT press.

Nyabola, N. (2018). *Digital democracy, analogue politics: How the internet era is transforming politics in kenya*. Zed Books.

Ochigame, R. (2019, Dec). How big tech manipulates academia to avoid regulation. Retrieved from https://theintercept.com/2019/12/20/mit-ethical-aiartificial-intelligence/

Pfotenhauer, S., & Jasanoff, S. (2017). Panacea or diagnosis? Imaginaries of innovation and the 'Mit model' in three political cultures. *Social Studies of Science, 47*(6), 783–810. https://doi.org/10.1177/0306312717706110

Potthast, T. (2014). *The values of biodiversity: philosophical considerations connecting theory and practice. Concepts and values in biodiversity*. Routledge.

Ranciere, J. (1998). *Disagreement: Politics and philosophy*. University of Minnesota Press.

Rijkhoff, J., Bakker, D., Hengeveld, K., & Kahrel, P. (1993). A method of language sampling. Studies in Language. *International Journal sponsored by the Foundation, 17*(1), 169–203.

Saad-Sulonen, J., Eriksson, E., Halskov, K., Karasti, H., & Vines, J. (2018). Unfolding participation over time: Temporal lenses in participatory design. *CoDesign, 14*(1), 4–16. https://doi.org/10.1080/15710882.2018.1426773

Schwartz, L. (2022, May). Primum Non Nocere: Before working with Indigenous data, the ACL must confront ongoing colonialism. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (vol. 2: Short papers) (pp. 724–731). Dublin, Ireland: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2022.acl-short.82 10.18653/v1/2022.acl-short.82

Schwemmer, C., Knight, C., Bello-Pardo, E. D., Oklobdzija, S., Schoonvelde, M., & Lockhart, J. W. (2020). Diagnosing gender bias in image recognition systems. *Socius*. https://doi.org/10.1177/2378023120967171

Sennrich, R., Haddow, B., Birch, A. (2015). Neural machine translation of rare words with subword units. arXiv preprint arXiv:1508.07909.

Smith, R.C., Winschiers-Theophilus, H., Loi, D., de Paula, R.A., Kambunga, A.P., Samuel, M.M., Zaman, T. (2021). Decolonizing design practices: Towards pluriversality. Extended Abstracts of the 2021 Chi Conference on Human Factors in Computing Systems. Association for Computing Machinery. Retrieved from https://doi.org/10.1145/3411763.3441334

Spivak, G. C. (1988). Can the subaltern speak. In L. Grossberg & C. Nelson (Eds.), *Marxism and the interpretation of culture* (pp. 66–111). University of Illinois Press.

Taylor, L., & Broeders, D. (2015). August). In the name of Development: Power, profit and the datafication of the global South. *Geoforum, 64*, 229–237. https://doi.org/10.1016/j.geoforum.2015.07.002

Thiong'o, N. w. (1986). Decolonising the mind: The politics of language in african literature. N.H: Heinemann, Oxford.

Tsing, A. L. (2012). On nonscalability: The living world is not amenable to precision-nested scales. *Common Knowledge, 18*(3), 505–524. https://doi.org/10.1215/0961754X-1630424

Vanmassenhove, E., Shterionov, D., Gwilliam, M. (2021, April). Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main volume (pp. 2203– 2213). Online: Association for Computational Linguistics. Retrieved from https://aclanthology.org/2021.eacl-main.188 10.18653/v1/2021.eacl-main.188

White, J.C., & Cotterell, R. (2021, August). Examining the inductive bias of neural language models with artificial languages. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (vol. 1: Long papers) (pp. 454–463). Online: Association for Computational Linguistics Retrieved from https://aclanthology.org/2021.acl-long.38 10.18653/v1/2021.acl-long.38

Winner, L. (1988). *The whale and the reactor: A search for limits in an age of high technology (Reprint* (Edition). University of Chicago Press.

Young, H. (2015). The digital language divide. Retrieved from https://labs.theguardian.com/digital-language-divide/

Young, I. M. (1990). *Justice and the politics of difference*. Princeton University Press.

Zaugg, I.A., Hossain, A., Molloy, B. (2022, Apr). Digitally-disadvantaged languages. Internet Policy Review, 11(2). Retrieved from https://policyreview.info/glossary/digitally-disadvantaged-languages 10.14763/2022.2.1654

Zevallos, R., & Bel, N. (2023). Hints on the data for language modeling of synthetic languages with transformers. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (vol. 1: Long papers) (pp. 12508–12522).

Zouhar, V., Chang, K., Cui, C., Carlson, N., Robinson, N., Sachan, M., Mortensen, D. (2023). Pwesuite: Phonetic word embeddings and tasks they facilitate. arXiv preprint arXiv:2304.02541.