
Technocultural Hegemony: What Role Does Natural Language Processing Play in the Reinforcement of Dominant Cultural Narratives?

Daria Markava
daria.markava@bist.ge

Abstract

While Natural Language Processing (NLP) tools keep gaining popularity among users from around the globe, their vast majority is developed in the west, mainly the US. Although plenty of studies have shown that NLP tools don't perform equally well in different languages and cultural contexts, little research has been conducted on the broader consequences of such performance disparities. By using the evidence from previous research, this study aims to bridge this gap and explore how NLP and the existing cultural hierarchies can be mutually constitutive. This paper first reviews existing literature on the NLP tools' performance in relation to underrepresented languages and non-western cultures. It then takes a critical theory approach to examining the broader cultural implications of the shortcomings identified during the review. More specifically, this work uses the concept of technoculture proposed by Leila Green, to connect the technological and cultural aspects of NLP, and refers to Gramsci's theory of cultural hegemony to explore how the bias in NLP tools reinforces dominant cultural narratives overrepresented in the training data. This study argues that NLP applications play a role in reinforcing the dominant norms, ideals, and ways of expression as universal, thus marginalizing alternative worldviews and imposing normative standards of communication onto the users of different backgrounds. The analysis concludes that, as the popularity of NLP tools keeps growing worldwide, their influence on what is perceived as "common sense" will increase too. This study emphasizes the importance of ensuring equitable representation of the user base throughout the whole NLP development pipeline.

I. Introduction

Although Natural Language Processing (NLP) has been around since the middle of the 20th century, it started gaining mass popularity outside the AI industry around 2017, sparked by Google's introduction of the Transformer architecture (Vaswani et al., 2017), and later by OpenAI's creation of GPT-1 (Radford et al., 2018). Increasingly, users from across the globe are using Large Language Models (LLMs) for various purposes on a regular basis, and the integration of these tools into everyday life is only expected to continue.

In contrast to the diversity of their user base, the vast majority of NLP tools (as well as other applications of artificial intelligence) are developed in the West - predominantly the United States - with China close behind. While a lot of those tools are nominally supported in many different countries and languages, there are barriers preventing people of different cultures from benefiting from the technology equally. These barriers, which will be discussed below in more detail, effectively disadvantage users with cultural backgrounds different from those of the developers, and perpetuate existing inequalities.

LLMs are unlike any other type of communication tools, as they don't enable interactions between people directly. Nevertheless, the text produced by LLMs is influenced by the real-world training data, meaning that although the final output is machine-generated, human influence is not eliminated from the pipeline, but rather takes a new form. In order to characterize the unique ways in which AI developers' cultural narratives are passed onto the users through NLP, this paper looks into the concepts of technoculture and cultural hegemony, and how they intertwine in NLP applications.

The purpose of this study is to take an interdisciplinary approach to identifying and understanding the implications of NLP tools, like LLMs, for users of diverse cultural backgrounds. While previous research on cultural aspects of NLP mentions the negative effects of its performance disparities across different languages and cultures, these effects are rarely discussed in depth, and the majority of studies focus solely on the technical aspects. There is currently little research on how the increased global use of NLP tools is going to reshape the culture worldwide, and this paper aims to address this gap by placing the effects of natural language processing on cultures around the world in a broader context.

II. Methodology

Research Questions

The key question this research aims to answer is how NLP reinforcing dominant cultural narratives contributes to cultural hegemony, epistemic injustice, and perpetuates existing

inequalities. To evaluate the cultural impact of NLP, this study will answer the following questions:

RQ1. To what extent can NLP be considered an emerging form of technoculture according to Green's definition?

RQ2. How do existing cultural hierarchies influence the performance of NLP tools?

RQ3. What could be the consequences of discrepancies of NLP applications' capabilities across different languages and cultural contexts?

Answering these questions will allow for a comprehensive understanding of how NLP can influence cultures, languages, and the relationships between them.

Literature Collection

The search for relevant literature was conducted across three databases: IEEE Xplore, ACM Digital Library, and Google Scholar. IEEE Xplore and ACM DL were chosen for their renown in the disciplines of technology and computer science; Google Scholar was used primarily due to its scope. Furthermore, these databases include literature produced by authors from different cultures around the world. This was a factor when choosing databases, since incorporating diverse perspectives is essential to reflecting and upholding the values that motivated this study in the first place.

To collect relevant literature, keywords "NLP" and "LLMs" were searched in combination with terms such as "culture", "narrative", "hegemony", "westernization", "dominance", "bias", etc., as well as boolean operators AND and OR. In all three databases, the advanced search option was used to search for keywords in either the title of an article or the full text. In the case of Xplore, publication topic filters were used as well. For terms like "culture", "dominance" and "epistemology", an asterisk was used to allow for relevant derived forms of the words to be included. While searching ACM Digital Library and IEEE Xplore, more general keyword combinations were used. The scope of Google Scholar, on the other hand, allowed for more precise queries, such as 'NLP AND "anglocentrism"', whereas similar queries on the other two databases returned no results. For all queries, the time range was specified as 2020-2025 to ensure the relevance of the articles, with the exception of the query 'NLP AND "epistemic injustice"', where the time range was set as 2024-2025, to limit the number of search results to a manageable amount. In ACM and Xplore, only articles labeled as "Open Access" or "Free" were selected; since Google Scholar doesn't provide such an option, 7 out of initially selected articles could not be accessed and were therefore excluded from screening.

Table 1 provides a summary of the search, selection and inclusion results.

N=	IEEE Xplore	ACM DL	Google Scholar	Total
Search results	163	116	468	747
Selected	23	8	54	85
Screened	23	8	47	78
Included	3	2	15	20

Table 1: Number of articles (N) yielded, selected, screened and included in the review.

Inclusion/exclusion criteria

To collect literature that would comprehensively represent the current state of NLP capabilities in relation to culture, the following inclusion and exclusion criteria were implemented. Included were studies that examined the performance of NLP tools in the context of language and/or culture. As expected, the majority of studies focused on LLMs; however, a few studies of Sentiment Analysis tools were also included. For a study to be included in the review, it had to discuss the question of culture, identity, fairness, or inclusivity raised by the study's results. Therefore, studies that were solely focused on the technical aspects of NLP applications were excluded. Works that proposed domain-specific applications of NLP (e.g. in healthcare, e-commerce, etc.) and review articles were excluded. Recognizing the rapid pace of development in the AI field, relevant preprints from arXiv found during the search of Google Scholar were included as well.

Critical theory analysis

Drawing from the evidence from literature, this work explores the cultural implications of NLP through the lens of two theories, technoculture and cultural hegemony.

Technoculture is a neologism that can be vaguely defined as the interaction between technology and culture. However, this definition lacks specificity, as it allows for any overlap between technology and culture to be considered technoculture (Green, 2001). To avoid using the term loosely, this study uses the definition of technoculture suggested by Green. Her paper defines technoculture as “communication of cultural material in technological contexts” that “can influence perceptions and experiences of time, place and space” (p. 16).

This research assesses to what extent NLP technologies fit Green's definition of technoculture. However, it is essential to keep in mind that the framework was developed in 2001, long before the NLP in its modern form was created, let alone widely employed by the general public. Still, Green acknowledged that new forms of technoculture would appear with time, which opens the

opportunity for investigating the role artificial intelligence, namely NLP, plays in the construction of culture.

The second concept this paper explores is cultural hegemony. Similarly to technoculture, there is no universally agreed-upon definition of the term, and its exact meaning depends on the sociocultural context to which it is applied. The theory of cultural hegemony was coined by an Italian Marxist philosopher Antonio Gramsci. It refers to the way the ruling class reinforces its dominance in society by positioning its values and beliefs as the norm via cultural institutions (Lears, 1985). Although the term was originally used to explain how the ruling elites maintain their power over the working class in capitalist societies, it has become more widely used in other contexts. The ones most relevant to this work are the linguistic hegemony of the English language worldwide, and the cultural aspects of Western hegemony, i.e. the positioning of Western ideals as universal. This article aims to determine to what extent these two forms of cultural hegemony manifest in modern NLP applications, with a focus on large language models.

III. NLP as a Technocultural Phenomenon

Green (2001) suggests that an ideal example of technoculture “would necessarily be in the high-tech domain”, which modern NLP applications undoubtedly are, given that they utilize state-of-the-art deep learning techniques. Green also argues that such an instance of technoculture would “be imbued with communication and culture”, and enable the creation and circulation of cultural material “using technologically mediated communication” (p. 16). Creation and circulation of cultural material is inherent to NLP for the following reasons.

First, these models are trained on enormous corpora of human-produced texts, such as books, web pages, online publications, social media posts, etc. These sources of text data are not neutral representations of language, but direct products of culture, permeated with human values, biases, and worldviews. Using them as training data results in the same epistemes being woven into NLP models. Consequently, the output of such tools is heavily influenced by the narratives present in the corpora, enabling the circulation of cultural material. An example of this can be seen in Das et al. (2024), who conducted an audit of Bengali sentiment analysis tools and found that most of the tools tested were biased towards particular genders, nationalities, and religions. This illustrates how post-colonial prejudices embedded in the training data are communicated through an NLP tool (in this case, by assigning more positive or negative sentiment scores to sentences expressing some or other identities).

Second, some types of NLP like natural language generation or machine translation produce language-based output, which in itself can be seen as a form of creation of cultural material since language, according to Green, is a technoculture as well. Additionally, by producing text output, generative language models create a body of text that is informed by, and builds on, human knowledge and culture. For example, when prompted to write an email to a professor, an LLM

would produce a formal and professional-sounding message, even if not explicitly asked to make it formal¹. Such an output is influenced by the norms of communication in settings such as academia and workplace; in its turn, the output may reinforce the user's understanding of these norms. Machine translation tools facilitate communication of culture more directly, by enabling the exchange of ideas between languages; however, they sometimes fail to communicate some linguistic nuances successfully (Helm et al. 2023).

Another characteristic of a technoculture, according to Green, is its ability to negate space or time (p. 12). In NLP, this ability manifests in quite an interesting way. As Johnson et al. (2022) point out, the training data for language models only represents the current moment in time. Contrary to the human culture, which evolves continuously, an individual NLP tool's model of reality is static, and doesn't take into account the changes that have taken place over time. Although modern NLP applications haven't been around long enough for this effect to be substantially visible, it is likely that a model released in 2025 has been trained on a different representation of reality than a model released in 2018.

As of June 2025, language models do not learn continuously in real time; once they have been trained, they reflect the cultural narratives of the time of their training for as long as they are used. A suitable analogy is a book (which Green also uses as an example of technoculture); a book written in the 19th century will continue to reflect the 19th century's values regardless of when it is being read. It is worth noting, however, that the time-negating properties of NLP might not be similarly obvious. New models are developed and fine-tuned at a rapid pace, replacing older versions; the "lifetime" of any language model, therefore, is incomparable to the period of time it would take for a significant shift of values to take place in society.

Overall, the above analysis suggests that NLP can indeed be considered an emerging form of technoculture, as it is simultaneously technological and rooted in human culture, with these two aspects being inseparable from each other.

IV. Cultural Hegemony in NLP

Building on Gramsci's theory of cultural hegemony (Lears, 1985), this section examines how the technocultural occurrence of NLP serves to reinforce dominant cultural narratives, particularly the westernization of cultural defaults globally and the supremacy of the English language as the universal lingua franca.

Cultural competence of NLP tools, mainly LLMs, has been widely studied (Aksoy, 2024; Dusi et al., 2024; Li et al., 2024; Naous et al., 2024; Pawar et al., 2025; Plaza-del-Arco et al., 2024; Shen et al., 2024; Tao et al., 2024; Wang et al., 2025), proving that the tools tested display some sort of bias based on factors such as gender, religion, nationality, race and other identity dimensions.

¹ An example of the prompt and output can be found here: <https://chatgpt.com/share/680de0ff-8e6c-8013-861b-2cf3385bfb4e>

NLP tools can be applied in various cases, such as content moderation, customer feedback analysis, social media monitoring, etc. (Shehu et al., 2024). When the tools used for such purposes exhibit bias based on demographic factors, it increases the risk of discrimination against the users whose cultural values, ways of expression, and linguistic backgrounds differ from those most prevalent in the training corpora.

Perhaps the largest cause of the unequal cultural representation in NLP tools is the training data imbalance. While there are massive English-language corpora available, the majority of the world's languages are underrepresented in the training datasets. For instance, on Common Crawl - a popular source of training data for NLP models - about 45% of the data is in English, as shown in Figure 1 (Dawson et al., 2024). Consequently, there are more opportunities for the development of English NLP than most other languages.

Language	CC-MAIN-2024-22 (%)	CC-MAIN-2024-26 (%)
English	45.9693	45.0546
Russian	5.9154	5.6876
German	5.5710	5.4701
Japanese	4.9833	4.7772
Spanish	4.6177	4.5025
Chinese	4.0651	4.8848
French	4.2157	4.2175
Italian	2.4205	2.6379
Portuguese	2.0407	2.0232
Dutch	1.8350	1.8288
Polish	1.7290	1.6771
Turkish	1.1142	1.1946
Malayalam	0.0226	0.0231
Yoruba	0.008	0.008

Figure 1: AI training data on Common Crawl across language groups (Dawson et al., 2024)

The availability of pre-trained models is also an issue for underrepresented languages. While training a language model from scratch requires a lot of resources, fine-tuning a pre-trained model is much less costly - and since these only exist for a few languages, the speakers of the remaining languages are further excluded from participation (Wali et al. 2020).

As a result, innovations in NLP are almost always centered around the English language. As Wali et al. observe, new tools are first introduced for English, only later adding support for other languages. This gives the English-speaking world a significant advantage, as they are often the first to get access to latest NLP technologies, and exacerbates the digital divide. Those who do not speak English or some other high-resource language are essentially excluded from the newest developments in NLP, and hence the benefits that the technology brings.

Furthermore, even when multiple languages are nominally supported by an NLP tool, it doesn't always mean that the speakers of those languages receive equally high quality services. Wali et al. emphasize that when low-resource languages are supported by a multilingual model, the model's performance in these languages tends to be significantly worse. While it is technically possible for the speakers of such languages to use NLP tools, they are still disadvantaged by this disparity.

This argument is also strongly supported by Helm et al. (2023), who define language modelling bias, a form of linguistic bias, as a situation when

“[T]he 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.” (p. 6)

This stems from the oversimplified idea of diversity in the field of technology, the study argues, and the work on expanding NLP to support more languages being usually done by the western developers, rather than the native speakers - a “top-down” approach that ultimately harms the native communities by forcing them to fit into the Anglocentric standards of communication in order to use the technology.

Such an approach creates an illusion of representation while in reality, it perpetuates marginalization of expression that doesn’t meet western standards, while positioning the latter as universal. Helm et al. uses Fricker’s concept of epistemic injustice to examine the harms caused by language technologies when they fail to adequately represent the voices of certain communities. Hermeneutical injustice, a type of epistemic injustice, refers to the way someone’s experience is not adequately understood or represented since the means of expressing it don’t exist (Fricker, 2007). This is true for NLP tools, especially LLMs, where worldviews and narratives alternative to the ones of the western world are hardly acknowledged. Lears (1985), quoting Gramsci, states that “every language contains the elements of a conception of the world” (p. 569). Thus unable to communicate the realities of their world when interacting with language models, speakers of underrepresented languages are forced to adapt to the privileged epistemes, reinforcing the notion that the dominant narratives are the only legitimate ones.

V. Conclusion

This study has shown how NLP possesses the key characteristics of a technoculture, particularly by synthesizing and enabling the circulation of cultural material, and negating the effects of time. A deeper analysis would be desirable; however, it would be beyond the scope of this paper. This work has also looked into the hegemonic elements of NLP, including how structural inequalities simultaneously influence and are held up by the technology; its tendency to misrepresent, stereotype and discriminate against many identities and demographic groups; and its positioning of western standards as the default. NLP, acting as a technoculture, perpetuates existing imbalances and reinforces dominant cultural narratives. As NLP tools’ usage increases in various applications, their hegemonic effects will manifest at a bigger scale, amplifying the voices of some at the expense of others’ exclusion.

References

- Aksoy, M. (2024). *Whose Morality Do They Speak? Unraveling Cultural Bias in Multilingual Language Models* (No. arXiv:2412.18863). arXiv. <https://doi.org/10.48550/arXiv.2412.18863>
- Das, D., Guha, S., Brubaker, J. R., & Semaan, B. (2024). The ``Colonial Impulse" of Natural Language Processing: An Audit of Bengali Sentiment Analysis Tools and Their Identity-based Biases. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–18. <https://doi.org/10.1145/3613904.3642669>
- Dawson, F., Mosunmola, Z., Pocker, S., Dandekar, R. A., Dandekar, R., & Panat, S. (2024). *Evaluating Cultural Awareness of LLMs for Yoruba, Malayalam, and English* (No. arXiv:2410.01811). arXiv. <https://doi.org/10.48550/arXiv.2410.01811>
- Dusi, M., Arici, N., Emilio Gerevini, A., Putelli, L., & Serina, I. (2024). Discrimination Bias Detection Through Categorical Association in Pre-Trained Language Models. *IEEE Access*, 12, 162651–162667. <https://doi.org/10.1109/ACCESS.2024.3482010>
- Etxaniz, J., Azkune, G., Soroa, A., de Lacalle, O. L., & Artetxe, M. (2024). *BERTAQA: How Much Do Language Models Know About Local Culture?*
- Fricker, M. (2007). *Epistemic Injustice*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198237907.001.0001>
- Green, L. (2001). Technoculture: Another Term That Means Nothing and Gets Us Nowhere? *Media International Australia*, 98(1), 11–25. <https://doi.org/10.1177/1329878X0109800105>
- Helm, P., Bella, G., Koch, G., & Giunchiglia, F. (2024). Diversity and language technology: How language modeling bias causes epistemic injustice. *Ethics and Information Technology*, 26(1), 8. <https://doi.org/10.1007/s10676-023-09742-6>
- Johnson, R. L., Pistilli, G., Menéndez-González, N., Duran, L. D. D., Panai, E., Kalpokiene, J., & Bertulfo, D. J. (2022). *The Ghost in the Machine has an American accent: Value conflict in GPT-3*.
- Lears, T. J. J. (1985). *The Concept of Cultural Hegemony: Problems and Possibilities*. *The American Historical Review*, Vol. 90, No. 3 (Jun., 1985), pp. 567-593
- Li, H., Jiang, L., Hwang, J. D., Kim, H., Santy, S., Sorensen, T., Lin, B. Y., Dziri, N., Ren, X., & Choi, Y. (2024). *CULTURE-GEN: Revealing Global Cultural Perception in Language Models through Natural Language Prompting* (No. arXiv:2404.10199). arXiv. <https://doi.org/10.48550/arXiv.2404.10199>
- Naous, T., Ryan, M. J., Ritter, A., & Xu, W. (2024). *Having Beer after Prayer? Measuring Cultural Bias in Large Language Models* (No. arXiv:2305.14456). arXiv. <https://doi.org/10.48550/arXiv.2305.14456>

- Nasution, A. H., & Onan, A. (2024). ChatGPT Label: Comparing the Quality of Human-Generated and LLM-Generated Annotations in Low-Resource Language NLP Tasks. *IEEE Access*, 12, 71876–71900. <https://doi.org/10.1109/ACCESS.2024.3402809>
- Pawar, S., Arora, A., Kaffee, L.-A., & Augenstein, I. (2025). *Presumed Cultural Identity: How Names Shape LLM Responses* (No. arXiv:2502.11995). arXiv. <https://doi.org/10.48550/arXiv.2502.11995>
- Plaza-del-Arco, F. M., Curry, A. C., Paoli, S., Curry, A., & Hovy, D. (2024). *Divine LLaMAs: Bias, Stereotypes, Stigmatization, and Emotion Representation of Religion in Large Language Models* (No. arXiv:2407.06908). arXiv. <https://doi.org/10.48550/arXiv.2407.06908>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). *Improving Language Understanding by Generative Pre-Training*.
- Rice, K. (2025). ChatGPT and World History Essays: An Assignment and its Insights into the Coloniality of Generative AI. *Teaching History: A Journal of Methods*, 49(1), 49–55. <https://doi.org/10.33043/67bdga4z9>
- Shehu, H. A., Usman Majikumna, K., Bashir Suleiman, A., Luka, S., Sharif, Md. H., Ramadan, R. A., & Kusetogullari, H. (2024). Unveiling Sentiments: A Deep Dive Into Sentiment Analysis for Low-Resource Languages—A Case Study on Hausa Texts. *IEEE Access*, 12, 98900–98916. <https://doi.org/10.1109/ACCESS.2024.3427416>
- Shen, S., Logeswaran, L., Lee, M., Lee, H., Poria, S., & Mihalcea, R. (2024). *Understanding the Capabilities and Limitations of Large Language Models for Cultural Commonsense* (No. arXiv:2405.04655). arXiv. <https://doi.org/10.48550/arXiv.2405.04655>
- Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), pgae346. <https://doi.org/10.1093/pnasnexus/pgae346>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is All you Need*.
- Wali, E., Chen, Y., Mahoney, C. M., Middleton, T. G., Babaeianjelodar, M., & Matthews, J. N. (2020). *Is Machine Learning Speaking my Language? A Critical Look at the NLP-Pipeline Across 8 Human Languages*.
- Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). *Large language models that replace human participants can harmfully misportray and flatten identity groups* (No. arXiv:2402.01908). arXiv. <https://doi.org/10.48550/arXiv.2402.01908>
- Yakhni, S., & Chehab, A. (2025). *Can LLMs Translate Cultural Nuance in Dialects? A Case Study on Lebanese Arabic*.
- Yang, E., & Roberts, M. E. (2021). Censorship of Online Encyclopedias: Implications for NLP Models. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 537–548. <https://doi.org/10.1145/3442188.3445916>