The Academic
International Journal of Multidisciplinary Research

An Online Peer Reviewed / Refereed Journal
Volume 4 | Issue 10 | October 2026
ISSN: 2583-973X (Online), SJIF IF(2026): 5.6
Website: www.theacademic.in

Whose English Does AI Speak? Bias, Standardization, and the Marginalization of Non-Native Varieties in Large Language Models

Nithyasree N

PRT English, The Guardian English School, Periya , Keralam

nithyasreeneelamana@gmail.com

DOI : https://doi.org/10.5281/zenodo.24394835

ARTICLE DETAILS

Research Paper

Received: 27-10-2026

Accepted: 02-11-2026

Published: 20-11-2026

Keywords:

large language models, World Englishes, linguistic bias, standard language ideology, dialect discrimination, algorithmic fairness, Indian English

ABSTRACT

Large language models (LLMs) increasingly mediate writing, learning, hiring, and scholarly communication across the English-using world, yet most of that world does not speak the variety these systems treat as normal. This article asks whose English LLMs recognize, reward, and reproduce, and what follows for speakers of Indian, Nigerian, Singaporean, African American, and other non-dominant Englishes. Through a critical integrative review of peer-reviewed research in natural language processing, human–computer interaction, and sociolinguistics, it connects World Englishes theory, standard language ideology, and raciolinguistic perspectives with recent empirical audits. The synthesis shows that disadvantage accumulates across five linked stages, termed the Standardization Cascade: corpus curation, model representation, preference alignment, user interaction, and downstream gatekeeping. Evidence includes comprehension gaps for non-standard varieties, covert dialect prejudice that survives alignment, AI-text detectors that misclassify non-native writing, and writing assistants that pull users toward Western norms. The paper argues that these effects amount to a quiet linguistic recentralization that threatens the pluricentric gains of World Englishes scholarship, and it offers five testable propositions and recommendations for developers, institutions,

Disclaimer: The opinions, interpretations, conclusions, recommendations, analyses, and statements expressed in this research article are solely those of the author(s) and do not reflect the views, policies, or positions of the Publisher, Chief Editor, Editors, Editorial Board, Reviewers, or their affiliated institutions. The author(s) are solely responsible for ensuring the originality, authenticity, and integrity of the manuscript and for ensuring that it is free from plagiarism, AI-generated content, copyright infringement, data falsification, and any other form of academic misconduct.

and policymakers, with attention to India.

Introduction

When a teacher in Kerala asks a chatbot to tidy a circular that begins "Kindly do the needful," the reply usually comes back without the phrase. In a narrow sense the tool has made no mistake. It has returned English that resembles the English it saw most often in training and that its human raters preferred. The small edit, however, carries a large assumption: that one way of writing English is correct and that the Englishes of India, Nigeria, Singapore, or the Caribbean are departures to be repaired. Repeated across millions of daily interactions, such edits begin to function less like assistance and more like language policy.

English now has more second-language and foreign-language users than first-language users, and its center of gravity moved away from Britain and North America long ago (Bhatt, 2001; Jenkins, 2006). Sociolinguists spent decades showing that postcolonial varieties developed their own stable norms and deserve recognition as legitimate Englishes rather than learner errors (Kachru, 1976, 1985; Schneider, 2003). Large language models (LLMs), by contrast, learn mostly from web text dominated by a few national varieties, filtered by heuristics designed around those varieties, and tuned by raters whose sense of quality reflects particular educational traditions (Bender et al., 2021; Dodge et al., 2021; Ouyang et al., 2022). These systems may be quietly rebuilding the hierarchy of Englishes that applied linguistics worked to dismantle.

Evidence on this problem is growing but scattered. Computer scientists report performance gaps on dialect benchmarks (Ziems et al., 2022, 2023), social scientists document covert prejudice (Hofmann et al., 2024), and education researchers question AI-text detectors and writing aids (Agarwal et al., 2025; Liang et al., 2023). Each strand tends to treat its result as an isolated defect, even though broad surveys define social bias as disparate treatment or outcomes rooted in historical and structural power asymmetries (Gallegos et al., 2024). Blodgett et al. (2020) cautioned that bias research in natural language processing (NLP) often lacks a clear account of what harm is being measured and why it matters socially. This paper addresses that gap by asking which variety of English LLMs treat as the default, how that default arises, and what disadvantages follow for speakers of other varieties. It contributes a five-stage model, the Standardization Cascade, that links findings across disciplines, together with propositions for future research in Outer Circle contexts such as India.

A brief note on terminology is needed. The native/non-native distinction ranks speakers by birthplace rather than competence and treats Inner Circle norms as the yardstick for everyone (Jenkins,

2006). Many speakers of Indian or Nigerian English acquire the language in early childhood. We therefore speak mainly of *non-dominant* varieties and use "non-native" only where the cited research does. African American English (AAE) is included deliberately: although it is a first-language variety, the mechanisms that disadvantage its speakers are closely related to those affecting Outer Circle users, because what matters is distance from an institutionally privileged standard (Flores & Rosa, 2015).

Theoretical Background

World Englishes

Kachru's (1985) model of concentric circles placed the traditional bases of English (the Inner Circle) alongside postcolonial societies where English holds institutional functions (the Outer Circle) and countries where it is learned as a foreign language (the Expanding Circle). Its lasting argument was that Outer Circle varieties are norm-developing systems with their own grammar, vocabulary, and pragmatics. Schneider's (2003) dynamic model traced how such varieties move from dependence on external norms toward endonormative stabilization, an achievement closely tied to postcolonial identity. Indian English illustrates the point. Its documented features include progressive forms with stative verbs ("I am having two sisters"), invariant tags ("You are coming, isn't it?"), focus particles such as *only* and *itself*, and words such as *prepone* and *revert* in the sense of "reply" (Sailaja, 2009). These forms are conventional in Indian offices, newspapers, and classrooms. To a system that measures quality by similarity to Inner Circle text, they look like anomalies.

Standard Language Ideology

Milroy (2001) argued that standardization works by suppressing optional variation, and that once a standard is established its users come to see it as the language itself. Lippi-Green (2012) described standard language ideology as a bias toward an abstracted, idealized, homogeneous language that is imposed and maintained by dominant institutions. Two elements of this account matter here: the standard is an abstraction that nobody fully speaks, and institutions such as schools, publishers, and employers keep it in place. LLMs are becoming one of those institutions. They encode a statistical abstraction of written English, they are deployed by organizations with enormous reach, and they present their outputs as neutral. Phillipson (1992) and, more recently, scholars of decolonial computing (Mohamed et al., 2020) remind us that such centralization reflects material inequalities rather than the intrinsic merit of any variety.

Raciolinguistics and the Listening Machine

Flores and Rosa (2015) shifted attention from the speaker to the listener, showing that racialized speakers are often heard as deficient regardless of how closely they match the standard. The insight transfers to machines. A model infers identity from linguistic cues, and its inferences draw on associations learned from text produced in unequal societies. We use the phrase *listening machine* for an LLM acting as an evaluator of the language it receives, a role that matters greatly when models grade essays, screen job applications, or judge credibility.

Method

The study uses a critical integrative review design, which synthesizes research from different methodological traditions to build new conceptual frameworks rather than to pool effect sizes (Snyder, 2019; Torraco, 2005). Sources were identified through purposive and snowball searches of the ACL Anthology, the ACM Digital Library, and journal databases in linguistics and multidisciplinary science, using terms such as *dialect bias*, *World Englishes*, *language model*, and *non-native English*. Peer-reviewed journal articles and archival conference proceedings were included if they examined a language technology and reported evidence on differential treatment of English varieties or of writers grouped by language background. Findings were then coded by the point in a model's development or use at which disadvantage originates, and these codes were refined into the five stages below. The set of studies is illustrative of the main lines of evidence rather than exhaustive.

Findings: The Standardization Cascade

The synthesis shows that disadvantage arises at five connected stages. We call the sequence a cascade because each stage feeds the next, so that small early skews are amplified later. Table 1 summarizes the stages.

Table 1

The Standardization Cascade: Stages, Mechanisms, Harms, and Possible Interventions

Stage	Core mechanism	Typical harm to non-dominant varieties	Possible intervention

Stage	Core mechanism	Typical harm to non-dominant varieties	Possible intervention
1. Curation	Web-scale collection and quality or toxicity filtering calibrated to standard text	Under-representation; silent removal of variety-marked text	Variety-aware data audits; filter evaluation by variety
2. Representation	Statistical learning from skewed data	Comprehension gaps; covert stereotypes attached to varieties	Disaggregated benchmarks using natural text; matched-guise probing
3. Alignment	Human preference ranking and reward modeling	Local forms rated as less polished; overt bias hidden, covert bias retained	Diverse rater pools; guidelines recognizing local standards
4. Interaction	Default output style and writing suggestions	Standardized replies; drift of users' writing toward Western norms	User-selectable variety settings; preservation of user voice
5. Gatekeeping	Model outputs used to score, screen, or detect	False misconduct flags; lower scores; exclusion from opportunities	Human review; ban on detector-only sanctions; fairness audits in procurement

Note. The cascade is cumulative: outputs of later stages, once published online, may re-enter future training corpora, reinforcing Stage 1.

Stage 1: Curation

Training data set the limits of what a model can learn. Bender et al. (2021) observed that the sheer size of web corpora does not guarantee diversity, because online participation is uneven. At the level of the world's languages, Joshi et al. (2020) showed that a handful of languages receive most data and tools while most others are barely represented; a similar skew operates among varieties of English. Everyday writing in Outer Circle varieties often circulates in spaces that are hard to crawl, such as messaging apps and regional media. Filtering deepens the gap. Dodge et al. (2021) found that the blocklist used to clean the Colossal

Clean Crawled Corpus disproportionately removed text in AAE and Hispanic-aligned English, and Blodgett et al. (2016) showed that language identification tools often misclassified AAE posts as not English at all. In each case, "clean" English is defined by proximity to a standard, so curation itself becomes standardization.

Stage 2: Representation

Trained models show uneven competence across varieties. Using Multi-VALUE, a framework covering 50 English dialects and 189 linguistic features, Ziems et al. (2023) found significant performance gaps for leading models on non-standard inputs, and they built gold-standard Indian English test data with native speakers. Deas et al. (2023) likewise found performance gaps between African American Language and White Mainstream English across six pretrained LLMs on generation tasks. Models also absorb social meanings. In a matched-guise study published in *Nature*, Hofmann et al. (2024) presented models with the same content in AAE and in Standard American English. The covert stereotypes attached to AAE speakers were more negative than any human stereotypes about African Americans recorded in experiments, and the models were more likely to assign these speakers less prestigious jobs and harsher criminal outcomes. Race was never mentioned; the variety alone triggered the prejudice.

Stage 3: Alignment

Assistants are refined through reinforcement learning from human feedback, in which raters rank outputs and the model is optimized toward their preferences (Ouyang et al., 2022). This aligns the model with particular people. The InstructGPT work relied on a team of roughly 40 contractors, whose preferences the authors acknowledged cannot represent all users. When raters favor concise, "professional" prose, localized politeness forms such as the Indian use of *kindly* may be steered away from as less polished. Alignment can also hide bias rather than remove it. Hofmann et al. (2024) reported that preference alignment widened the gap between overt and covert stereotypes, making openly prejudiced statements less likely while leaving dialect-linked prejudice beneath the surface. A polite model can still be a prejudiced one.

Stage 4: Interaction

Fleisig et al. (2024) prompted GPT-3.5 Turbo and GPT-4 with text from ten English varieties, including Indian, Nigerian, Kenyan, Jamaican, and Singaporean English. The models replied in standard varieties by default, and native-speaker evaluators rated responses to non-standard varieties as worse on stereotyping (19%), demeaning content (25%), comprehension (9%), and condescension (15%). GPT-4 improved comprehension but showed more stereotyping than its predecessor. Influence also runs from

model to user. In an experiment with 118 participants from India and the United States, Agarwal et al. (2025) found that AI writing suggestions gave Americans larger efficiency gains and led Indian participants to adopt Western writing styles. Each accepted suggestion is small and voluntary, which is why the cumulative drift is easy to overlook.

Stage 5: Gatekeeping

The final stage arises when model judgments are used to evaluate people. Liang et al. (2023) tested seven AI-text detectors on 91 TOEFL essays by non-native writers and 88 essays by American eighth-grade students. More than half of the non-native essays were wrongly flagged as AI-generated, while the native essays were classified almost perfectly. The detectors relied on perplexity, so plainer word choices typical of learners looked machine-made. This pattern extends a longer history: commercial speech recognition produced nearly twice the word error rate for Black speakers as for white speakers (Koenecke et al., 2020), and hate speech classifiers flagged AAE tweets as offensive more often (Sap et al., 2019). Markl (2022) documented similar disparities across regional and social accents of British English, showing that the problem is not confined to race or to the United States. As LLMs enter essay scoring, hiring, and moderation, linguistic difference can become material disadvantage.

How the Stages Compound

The stages reinforce one another. Curation thins a variety in the data, representation encodes it weakly and attaches surviving stereotypes, alignment rewards outputs that avoid it, interaction nudges users toward the standard, and gatekeeping judges users against that standard. A feedback loop then closes the cascade: AI-assisted text enters the web and may be collected for future training. If assisted writing in India gradually sheds local features, later corpora will contain even less of them. We call this tendency *regression to the standard*.

Discussion

The evidence supports a claim stronger than the familiar statement that AI contains bias. LLMs are becoming a standardizing institution in the sense described by Milroy (2001) and Lippi-Green (2012). Like dictionaries and examination boards, they define correctness, but with far greater reach and without public debate. The pluricentric settlement that World Englishes research helped secure is vulnerable to a quiet reversal.

India shows the stakes clearly. English there is a language of higher education, the courts, and the formal economy, taught largely by teachers who themselves speak Indian English. If writing tools consistently rewrite Indian usage, some users may come to view their own English as deficient, while others may learn to write for the machine rather than for their readers. In both cases, the authority to define good Indian English shifts toward developers located elsewhere. The smaller productivity gains observed among Indian users (Agarwal et al., 2025) also suggest that the benefits of these tools flow mainly to those who already speak the preferred variety.

Education and publishing face immediate risks. Detector-based misconduct procedures expose multilingual students to false accusations (Liang et al., 2023), and feedback tools that mark local forms as errors may undermine teaching that values students' full linguistic repertoires (Canagarajah, 2013). Researchers from the Global South increasingly use LLMs to polish manuscripts; this may lower one barrier while homogenizing scholarly voice. Journals should judge clarity, not conformity to one national variety.

The workplace raises related concerns. Employers increasingly use LLM-based tools to screen résumés, draft correspondence, and evaluate written tests. If a model reads Indian or Nigerian English as less competent, as the covert judgments reported by Hofmann et al. (2024) suggest it may, applicants could be filtered out before any human reads their work. Because such decisions are rarely explained, affected candidates have little chance to contest them. Organizations that adopt these tools in multilingual labor markets therefore carry a responsibility to test them on the varieties their applicants actually use, and to keep a human reviewer in the loop whenever language quality influences an outcome.

Propositions for Future Research

P1. The share of a variety removed during corpus filtering will rise with its distance from Standard American or British English.

P2. Model quality for a variety will be predicted more strongly by its share of training data than by its number of speakers.

P3. Preference alignment will reduce overt but not covert dialect bias, and this pattern will extend to Indian and Nigerian English.

P4. Sustained use of AI writing assistance will reduce local variety features in users' unassisted writing over time.

P5. AI-mediated assessment will penalize writers of non-dominant varieties more than human raters from the writers' own speech community.

Recommendations

Developers should report model performance separately for each English variety using naturally occurring text, test for covert bias with matched-guise designs, audit data filters for differential removal, include Outer Circle speakers among preference raters, and offer user-selectable variety settings that cover grammar and idiom, not only spelling. Universities and journals should stop treating detector scores as evidence of misconduct and should state that recognized local varieties are acceptable where meaning is clear. Policymakers in multilingual countries could require evidence of fair performance across locally spoken varieties before public bodies adopt AI systems. India's language technology programs, which rightly prioritize Indian languages, could extend that commitment to Indian English.

Limitations

As a review, this paper synthesizes rather than generates data, and its study set is not exhaustive. Much of the strongest evidence concerns AAE, so claims about Outer Circle varieties rest partly on extrapolation, which the propositions are designed to test. Several studies examined models that have since been updated. The structural mechanisms described here, however, are not tied to any single model version.

Conclusion

Current LLMs speak, by default, a written, educated, Inner Circle English, most often its American form, and they treat other Englishes as deviations to be accommodated, corrected, or penalized. This outcome does not stem from one design choice but from a cascade in which curation, representation, alignment, interaction, and gatekeeping each tilt toward the standard. Sociolinguistics established that Indian, Nigerian, Singaporean, African American, and other Englishes are systematic and legitimate. Whether that recognition survives will depend on whether developers, institutions, and speech communities insist that AI be judged by how well it serves the full range of English users, rather than by how closely those users can be made to resemble one group.

References

Agarwal, D., Naaman, M., & Vashistha, A. (2025). AI suggestions homogenize writing toward Western styles and diminish cultural nuances. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3706598.3713564

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bhatt, R. M. (2001). World Englishes. *Annual Review of Anthropology, 30*, 527–550. https://doi.org/10.1146/annurev.anthro.30.1.527

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). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485

Blodgett, S. L., Green, L., & O'Connor, B. (2016). Demographic dialectal variation in social media: A case study of African-American English. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 1119–1130). Association for Computational Linguistics. https://doi.org/10.18653/v1/D16-1120

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations*. Routledge. https://doi.org/10.4324/9780203073889

Deas, N., Grieser, J., Kleiner, S., Patton, D., Turcan, E., & McKeown, K. (2023). Evaluation of African American Language bias in natural language generation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 6805–6824). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.421

Dodge, J., Sap, M., Marasović, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., & Gardner, M. (2021). Documenting large webtext corpora: A case study on the Colossal Clean Crawled Corpus. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1286–1305). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.emnlp-main.98

Fleisig, E., Smith, G., Bossi, M., Rustagi, I., Yin, X., & Klein, D. (2024). Linguistic bias in ChatGPT: Language models reinforce dialect discrimination. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 13541–13564). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.750

Flores, N., & Rosa, J. (2015). Undoing appropriateness: Raciolinguistic ideologies and language diversity in education. *Harvard Educational Review, 85*(2), 149–171. https://doi.org/10.17763/0017-8055.85.2.149

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., Yu, T., Zhang, R., & Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics, 50*(3), 1097–1179. https://doi.org/10.1162/coli_a_00524

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). AI generates covertly racist decisions about people based on their dialect. *Nature, 633*(8028), 147–154. https://doi.org/10.1038/s41586-024-07856-5

Jenkins, J. (2006). Current perspectives on teaching World Englishes and English as a lingua franca. *TESOL Quarterly, 40*(1), 157–181. https://doi.org/10.2307/40264515

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 6282–6293). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.560

Kachru, B. B. (1976). Models of English for the Third World: White man's linguistic burden or language pragmatics? *TESOL Quarterly, 10*(2), 221–239. https://doi.org/10.2307/3585643

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.

Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J. R., Jurafsky, D., & Goel, S. (2020). Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences, 117*(14), 7684–7689. https://doi.org/10.1073/pnas.1915768117

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns, 4*(7), Article 100779. https://doi.org/10.1016/j.patter.2023.100779

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States* (2nd ed.). Routledge. https://doi.org/10.4324/9780203348802

Markl, N. (2022). Language variation and algorithmic bias: Understanding algorithmic bias in British English automatic speech recognition. In *Proceedings of the 2022 ACM Conference on Fairness,*

Accountability, and Transparency (pp. 521–534). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533117

Milroy, J. (2001). Language ideologies and the consequences of standardization. *Journal of Sociolinguistics, 5*(4), 530–555. https://doi.org/10.1111/1467-9481.00163

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology, 33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*, 27730–27744. https://doi.org/10.48550/arXiv.2203.02155

Phillipson, R. (1992). *Linguistic imperialism*. Oxford University Press.

Sailaja, P. (2009). *Indian English*. Edinburgh University Press.

Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019). The risk of racial bias in hate speech detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1668–1678). Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1163

Schneider, E. W. (2003). The dynamics of New Englishes: From identity construction to dialect birth. *Language, 79*(2), 233–281. https://doi.org/10.1353/lan.2003.0136

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human Resource Development Review, 4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Ziems, C., Chen, J., Harris, C., Anderson, J., & Yang, D. (2022). VALUE: Understanding dialect disparity in NLU. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 3701–3720). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.258

Ziems, C., Held, W., Yang, J., Dhamala, J., Gupta, R., & Yang, D. (2023). Multi-VALUE: A framework for cross-dialectal English NLP. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 744–768). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.44