

Linguistic Inequity in Large Language Models:

Assessing Bias, Multilingual Performance, and Dialect Erasure in ChatGPT

Kumar Shrestha

PhD Student, English Department, University of Cincinnati, Ohio

Email: shrestkr@mail.uc.edu

Abstract

This article presents a critical AI audit of ChatGPT, examining how its labor practices, data inputs, historical development, and outputs perpetuate deep-seated linguistic inequality. It situates the platform within a lineage shaped by Cold War military research, Silicon Valley corporatization, and English-dominated digital infrastructure. The audit demonstrates that ChatGPT exhibits unequal accuracy across languages, systematically normalizes non-standard dialects, and enforces homogenized stylistic norms. These practices result in the erasure of dialects, cultural flattening, and the reinforcement of global linguistic hierarchies tied to race, colonial history, class, and technological access. Consequently, speakers of marginalized languages face disproportionate barriers to information, recognition, and digital participation. The paper concludes by advocating for audit-based interventions to promote linguistic justice, including community-accountable deployment, reward structures that value linguistic diversity, and culturally grounded training data. It argues that treating language models as socially embedded technologies is essential for achieving equitable AI.

Keywords: AI ethics and audit, dialect erasure, large language models, language ideologies, multilingual bias

INTRODUCTION

ChatGPT, a generative artificial intelligence chatbot, was first made widely available outside of internal research use on November 30, 2022, when OpenAI formally released it to the public (ChatGPT, the Generative AI Chatbot, is released). Using the conversational capacities of the ChatGPT 3.5 architecture, ChatGPT was built on years of development in large language models.

As users engaged with the chatbot online, it rapidly gained international attention. With its quick update, ChatGPT became one of the fastest-growing consumer technologies in history and sparked a lot of conversation about the implications of AI for society going forward (ChatGPT, the Generative AI Chatbot, is released).

Large Language Models (LLMs) such as ChatGPT have become central to contemporary

Article History

Received: 18 Dec. 2025

Revised: 24 Dec. 2025

Accepted: 26 Dec. 2025

Cite as:

Shrestha, K. (2025). Linguistic inequity in large language models: Assessing bias, multilingual performance, and dialect erasure in ChatGPT. *NOU Multidisciplinary Journal*, 1(1) 99-109.

communication, education, research, and global technological interaction. Their research extends across linguistically diverse populations, yet these systems do not serve all languages and dialects equally. Several scholars have demonstrated that LLMs inherit the linguistic biases of their training data, thereby reinforcing global hierarchies that are shaped by factors such as race, colonial history, class, and technological access (Birhane 389; Bender et al. 615). This creates a self-perpetuating cycle in which high-resource, standardized languages, such as those primarily used in the U.S. and British contexts, receive robust computational representation, while marginalized, low-resource, and indigenous languages become increasingly invisible in digital spaces (Pakray et al., 2025).

About 800 million weekly active users are reportedly served by the platform as of late 2025, up from about 400 million earlier in the year (DemandSage). OpenAI reports that its recurring revenue from subscriptions, enterprise tools, and API services has reached \$10 billion annually (DemandSage, TechCrunch). ChatGPT is positioned as a dominant force in digital communication and knowledge production due to its extensive usage and financial clout. The enormous reach of such large language models also amplifies the potential repercussions of linguistic injustices ingrained in these systems. However, these systems are deeply embedded with the potential consequences of linguistic injustices. Because users who communicate in non-standard dialects, minority languages, or less-represented linguistic contexts may face discrimination, uneven multilingual performance, or even dialect erasure, there are significant issues with fairness, accessibility, and representation.

This AI audit examines ChatGPT as a case study in linguistic inequity. By analyzing the model's genealogy, data sources, labor practice, and output behavior, I argue that ChatGPT reproduces systematic patterns of dialect eraser, linguistic discrimination, and cultural flattening. These harms disproportionately impact communities that have been historically excluded from global knowledge infrastructures.

History of ChatGPT

The development of ChatGPT has its roots in a lengthy political and historical lineage influenced by Western linguistic ideologies, academic institutions, and military research. During the Cold War, machine translation projects focused on strategically important languages like Arabic, Chinese, and Russian were funded by U.S. defence agencies like the Department of Defence and DARPA. This led to the emergence of early computational linguistics, value was created by this military-academic system, which marginalized other languages while elevating standardized English and a select group of "geopolitically important" languages. Researchers like Bernadette Longo demonstrate how early computing was shaped by the confluence of military spending, national security priorities, and linguistic standardization in ways that still affect modern AI systems and their unequal treatment of languages.

These linguistic hierarchies were made worse by new economic and technological forces as AI development spread to Silicon Valley and corporate labs. Businesses like OpenAI expanded due to Microsoft's substantial investment and cloud infrastructure, which enabled the training of enormous models. Due to the concentration of AI power within English-dominated platforms, thousands of

African, Indigenous, Pacific, Himalayan, and Southern Asian languages remained underrepresented, and Euro-American digital corpora such as Wikipedia, news archives, books, and social media became the main sources of training data. Simultaneously, the growth of large language models depended on hidden global labor: under exploitative conditions, outsourced data annotators in Kenya, India, Venezuela, and the Philippines performed crucial labelling and moderation tasks. These methods of data extraction and labor reflect what academics like Couldry and Mejias refer to as “data colonialism,” showing how historical injustices persist in contemporary AI development.

Historical and Genealogical Analysis of ChatGPT

Any audit of an AI system must begin with its history. ChatGPT’s genealogy is shaped not only by military research, US federal funding, and Western linguistic standardization but also by the political economy of Silicon Valley’s major investors. These forces, such as military infrastructures, corporate capital, venture funding, and global data extraction, together determine which languages receive technological representation and which remain unseen.

Military and Academic Roots

ChatGPT-based models have their intellectual roots in computational linguistics during the Cold War, when U.S. military interest in Arabic, Chinese, and Russian stimulated machine translation research. These early systems received significant support from agencies, such as the U.S. Department of Defense and DARPA, embedding a geopolitical hierarchy of “computationally important” languages that continues to shape AI development today. This military-academic foundation created

the original divide between data-rich and data-poor languages, privileging linguistic communities considered strategies valuable. Bernadette Longo’s *Words and Power* demonstrates how Cold War politics, military spending, and nationalist ideals influenced early computer and language technologies in the United States. She explains that the need to process strategic languages like Russian led to the growth of machine translation and computational linguistics within a military-academic system backed by organizations like the Department of Defense. According to Logo, this time period influenced which linguistic knowledge was given priority by reinforcing the notion that standardized forms of English were better politically and technologically. By examining how Cold War values shape research agendas, she demonstrates how power and language standardization became embedded in computing in ways that still affect language technologies today.

Silicon Valley, Corporate Capital, and Big Tech Consolidation

As AI research shifted into corporate laboratories in the 2010s, new forms of linguistic hierarchy emerged. OpenAI, founded with early backing from Elon Musk, Sam Altman, Greg Brockman, Reid Hoffman, Peter Thiel, Jessica Livingston, etc., was initially funded through a network of Silicon Valley elites committed to accelerating AI capabilities. Venture capital firms like Khosla Ventures, Tiger Global, and Andreessen Horowitz also played influential roles in scaling the organization’s research.

The most transformative investment came from Microsoft, whose multibillion-dollar funding agreements (starting with \$ 1 billion in 2019 and expanded through 2023) (Microsoft, “Introducing OpenAI”) provided

the cloud infrastructure, supercomputing clusters, and Azure-based GPU resources needed to train large-scale computational power, further aligning AI development with English-dominant Western platforms. This consolidation reinforced English as the default language of the internet, while Euro-American text sources, such as Wikipedia, news corpora, digitalized books, and Reddit, became the backbone of model training (OpenAI, 2023). Meanwhile, African, Indigenous, Pacific, Himalayan, and many South Asian languages remained starkly underrepresented.

Hidden History of Data Extraction

The development of large language models depends not only on corporate investment but also on the hidden labor of data workers across the Global South. Investigations show that outsourced annotators in Kenya, the Philippines, Venezuela, and India performed toxicity labeling, linguistic categorization, and reinforcement learning tasks for wages as low as \$1.32 per hour (Perrigo, Exclusive; Perrigo, AI's Exploited Workforce). These labor conditions illustrate what Couldry and Mejias conceptualize as data colonialism, in which the extraction of human knowledge mirrors earlier resource-extraction economies (Couldry & Mejias, 2019).

Data Inputs: Whose Language Counts?

Weissweiler et al.'s "Counting the bugs in ChatGPT's Wugs: A Multilingual Investigation into the Morphological Capabilities of a Large Language Model" reveals that the unequal distribution of linguistic data in ChatGPT's training corpus significantly shapes its language capabilities, raising important questions about which languages are given priority. According to Weissweiler et al., there are

significant differences in the underlying data that support ChatGPT's linguistic competency, as evidenced by the model's strong performance on English and other high-resource languages but its difficulties with the morphology of low-resource and typologically diverse languages.

Labor Inputs: Invisible Human Workers

A key component of ChatGPT's conversational capabilities, Reinforcement Learning from Human Feedback (RLHF), depends on large teams of human annotators whose work is mostly overlooked in popular accounts of artificial intelligence. A large portion of this work is outsourced to nations in the Global South, where annotators are paid poorly to perform tasks, such as evaluating linguistic accuracy, labelling biased content, and ranking outputs. According to the studies, these employees are frequently instructed to follow linguistic rules based on Western norms, reward standardized grammar, and penalize non-standard grammar (Gray & Suri, 2019). This creates a paradox wherein multilingual workers, many of whom speak English as a second language, are paid to perpetuate linguistic norms that minimize the variety of their speech patterns and the diversity of their own communities. These practices act as a kind of linguistic gatekeeping that directly includes structural bias into the model's behavior long before users interact with the system. As a result, the work that goes into creating ChatGPT supports to creation of a linguistic hierarchy where some varieties of English are given preference while others are systematically excluded.

Energy and Natural Resource Input

Large language models (LLMs) such as ChatGPT have significant environmental and ethical implications. Lifecycle analyses indicate that training even moderately sized

models can emit hundreds of tons of CO₂ and consume millions of liters of water, primarily due to energy-intensive hardware and data-center operations (Morrison et al., 2025). The energy demands of LLMs, often powered by fossil-fuel-based grids, disproportionately impact communities already vulnerable to climate and resource inequities (Hinov). These environmental costs are inseparable from the social and economic structures that support AI development, as wealthier regions benefit from access to advanced language technologies while resource extraction and energy consumption disproportionately burden less-developed areas (Hinov). Moreover, the vast datasets required for training LLMs often overrepresent dominant languages, marginalizing minority languages and reinforcing global linguistic hierarchies (Morrison et al., 2025). Addressing these challenges requires integrating environmental sustainability into AI governance, including energy-efficient model design, renewable-powered data centers, and careful consideration of the global distribution of costs and benefits (Hinov). Ultimately, achieving linguistic justice in AI is inseparable from promoting ecological responsibility and equitable access, ensuring that technological advancement does not exacerbate social and environmental disparities (Morrison et al., 2025)).

Outputs: What ChatGPT Produces

Understanding what ChatGPT produces is essential for assessing its broader social impact, especially in multilingual and multicultural contexts. Although the model is often presented as neutral and universally capable, its outputs reveal patterns that mirror global linguistic hierarchies. ChatGPT is trained on disproportionately large amounts of English and other high-resource languages.

Its performance tends to privilege dominant linguistic groups while marginalizing speakers of underrepresented languages or dialects. These disparities affect not only the accuracy of response but also how the model interprets the tone, politeness, and identity. Consequently, the outputs of ChatGPT can reinforce existing inequities by elevating certain varieties as “Standard” or “Correct,” while subtly devaluing others. This section discusses three interesting forms of linguistic inequalities observed in ChatGPT’s outputs: unequal performance across languages, the erasure of non-standard dialects, and the homonization of tone and politeness norms.

Unequal Performance Across Languages and Nationality Bias

ChatGPT outputs reveal consistent disparities in accuracy across languages. It performs best in high-resourced languages, especially U.S. and British English, followed by major European languages, because these languages dominate the digital corpora used for training. Malik et al., in this article, “Language of discrepancies in the performance of generative artificial intelligence models: An Examination of Infectious Disease Queries in English and Arabic,” argue that ChatGPT’s responses in English were significantly more complete, accurate, and relevant than its responses in Arabic, revealing a structural imbalance rooted in the disproportionate availability of high-quality English training data. The authors argue that such disparities carry real-world consequences, as they may limit access to trustworthy medical information for millions of Arabic speakers and other communities using low-resource languages.

Dentella et al. (2023) highlight that current large language models, including ChatGPT, show low language accuracy and inconsistent

responses across repeated prompts, undermining their reliability in grammatical judgements (Dentella et al., 2023). The model also exhibits a “yes-response” bias, answering affirmatively regardless of grammatical correctness. These shortcomings highlight the risk of uneven and unreliable performance across different languages or linguistic contexts (Dentella et al., 2023). Shucheng Zhu et al., (2024) in this article, “Quite Good, but Not Enough: Nationality Bias in Large Language Models – A Case Study of ChatGPT,” argue that ChatGPT’s seemingly neutral and friendly responses across different nationalities can conceal underlying “nationality bias”. Their study, which tested the model across 195 countries using multiple prompts in both Chinese and English, demonstrates that certain prompts, particularly those with a negative undertone, can elicit stereotypical or biased descriptions associated with specific nationalities. Furthermore, the degree and nature of this bias vary depending on the language of the prompt and the cultural contexts, indicating that linguistic factors interact with and can amplify nationality-based prejudice present in the training data. The study underscores that, despite its surface-level reliability, ChatGPT is not immune to subtle but consequential biases, especially when addressing sensitive topics related to nationality.

Finally, these studies demonstrate that ChatGPT’s performance varies significantly across languages and nationalities. It performs best in high-resource languages like English, while low-resource languages such as Arabic exhibit lower accuracy and completeness, reflecting imbalances in training data (Sallam et al., 2025). The model also shows inconsistencies, response bias, and limitations in grammatical judgement (Dentella et al., 2023), and even seemingly

neutral outputs can hide subtle nationality biases influenced by language and cultural context (Zhu et al., 2024). These findings highlight the need for more inclusive and culturally aware AI training to ensure fair and reliable performance worldwide.

Dialect Erasure

Linguistic inequity is also evident within individual languages, where ChatGPT marginalizes non-standard dialects. When users input regional or socially stigmatized dialects, the models frequently normalize the text into Standard American or British English. Fleisig et al. (2024) argue that ChatGPT’s responses overwhelmingly defaulted to standardized forms, even when non-standard varieties were clearly used in the prompt.

Moreover, ChatGPT’s outputs frequently erase crucial features of linguistic identity, such as phonological variation, culturally specific lexical items, and community-based rhetorical styles. When input dialectal language, the model often produces standardized paraphrases that remove markers of racial or cultural identity, contributing to what scholars describe as dialect erasers (Fleisig et. al., 2024). This eraser matters because language is not only a tool for communication but a site of identity, history, and resistance. Therefore, linguistic outputs of ChatGPT reflect broader ideological forces: they reproduce dominant norms of correctness and intelligibility, while diminishing the visibility and legitimacy of marginalized linguistic communities.

Tone and Politeness Standardization

Large Language Models like ChatGPT often default to a tone of neutral friendliness and standardized politeness, a pattern that can diminish diversity in incommunicative

styles. Research demonstrates that emotional framing of a prompt, whether polite, neutral, or rude, can significantly influence ChatGPT's responses, leading the model to align its tone with user input (Bardol, 2025). This suggests that, even before grammar or lexical variation is considered, ChatGPT homogenizes its speech patterns, privileging a generic polite register over culturally specific expressions of directness, humor, or informality. Such standardization may flatten the richness of human speech and reduce the model's adaptability to diverse communicative norms across linguistic and cultural contexts.

Moreover, tone standardization intersects with cultural expectations of politeness, meaning that what counts as "polite" or "neutral" varies across communicative communities. Studies on prompt politeness reveal that the effectiveness and consistency of large language models vary depending on both tone and language, and that politeness strategies that function well in English do not necessarily transfer to other linguistic contexts (Yin et al., 2024). As a result, ChatGPT's default politeness bias may obscure culturally appropriate forms of directness, intimacy, or pragmatic style, effectively imposing a uniform tone that aligns with norms embedded in English-dominant training data. This creates a rhetorical hierarchy in which certain tone styles are implicitly positioned as more appropriate or "correct" than others, undermining linguistic pluralism.

Finally, the standardization of tone may have broader social and ethical implications. By consistently producing politeness, neutral, and "safe" outputs, ChatGPT risks erasing communicative markers of identity such as humor, sarcasm, regional friendliness, or culturally embedded rhetoric. When these linguistic signals are flattened, speakers from

diverse backgrounds may feel misrepresented or pressured to adjust their language to align with standardized norms to be understood or validated. Scholars note that such an eraser can reinforce power imbalances by privileging dominant linguistic norms over minority communicative practices (Dobariya & Kumar, 2025). In multilingual, multicultural, and educational contexts, tone standardization thus functions as a gatekeeping mechanism that determines whose voices and communicative styles are recognized as legitimate.

Consequences: Who Benefits and Who is Harmed?

The unequal design and output of large language models (LLMs) like ChatGPT produce tangible sociolinguistic consequences, disproportionately benefitting speakers of high-resource and standard language varieties while harming speakers of marginalized, underrepresented, or non-standard languages and dialects. Research supports this claim: for instance, "Linguistic Bias in ChatGPT: Language Models Reinforce Dialect Discrimination" (Fleising et al., 2024) demonstrates that when ChatGPT is prompted with non-standard English dialects (e.g., Nigerian, Indian, or Jamaican English), the model often normalizes them into American or British English. In such cases, the identity, voice, and communicative norms of speakers of non-standard dialects are systematically marginalized, reinforcing existing linguistic hierarchies rooted in race, class, and colonial histories.

Moreover, the harm extends beyond dialect within language to cross-linguistic inequities; speakers of low-resource languages often receive inferior performance from LLMs, undermining equitable access to information, services, and digital representation. Recent

work by Mind the Language Gap: Automated and Augmented Evaluation of Bias in LLMs for High- and Low-resource Languages (Buscemi et al., 2025) reveals persistent quality and bias disparity when LLMs handle queries in low-resource languages, despite the model's multilingual capacities. Similarly, the phenomenon of "stereotype leakage", documented in Multilingual Large Language Model Leak Human Stereotypes across Language boundaries (Cao et al., 2025), shows that harmful social biases, e.g., about nationality, ethnicity, or social group) learned one language can be transferred to others, meaning that multilingual LLMs may propagate prejudice across languages rather than mitigate it. Thus, in multilingual and cross-cultural contexts, LLMs may reinforce global inequalities of representation and dignity.

On a structural level, these inequities entrench what might be called a form of "digital linguistic colonialism." By systematically privileging standard, majority-language forms and penalizing non-standard or low-resource languages/dialects, LLMs contribute to a homogenization of global digital communication. This process yields benefits, such as convenience, fluency, and visibility to communities. The consequences include not only misrecognition and alienation, but also potential loss of linguistic diversity, cultural heritage, and equitable access to digital tools. What is often presented as a neutral tool for linguistic assistance instead becomes a mechanism that reproduces and intensifies preexisting inequities, rather than mitigating them.

Audit Intervention: Designing Inclusive and Culturally Grounded LLMs

I concentrate on how structural linguistic injustices are ingrained in the creation and

training of big models when analyzing the audit intervention for ChatGPT. My analysis emphasizes the need for approaches that prioritize inclusivity, fairness, and cultural sensitivity over technical efficiency. I contend that language models must be held accountable to linguistic justice principles by examining the causes of linguistic bias and taking into account interventions that diversify data, value linguistic variation, and encourage transparency. Instead of replicating prevailing norms, I believe that an equitable AI system should reflect the lived linguistic realities of marginalized and underrepresented communities.

Diversity Training Data

Many of the disparities in ChatGPT's performance, in my opinion, are caused by its heavy reliance on English-dominant corpora, which excessively favor standardized varieties and marginalize regional, indigenous, and non-standard forms. Increasing training data is, in my opinion, a critical first step in addressing these disparities. A model that is better able to respect linguistic identity rather than reduce it to a uniform norm when it is exposed to culturally grounded narratives, regional dialects, and common colloquialisms. This diversification, in my opinion, is a commitment to appreciating the languages and voices that have traditionally been disregarded rather than just a technical change.

Reward Linguistic Diversity in Model Training

Additionally, my analysis shows that linguistic inequality is reproduced in reinforcement learning protocols. Current reward systems unintentionally penalize speakers of marginalized dialects because they implicitly favor standardized grammar. I contend that dialect erasure and tone homoneziation could be avoided by rearranging these

reward systems to acknowledge and validate linguistic variation. In my opinion, rewarding such diversity promotes AI outputs that more accurately reflect users' identities and communication styles while also honoring the social and cultural significance of non-standard forms.

Transparent and Responsible Deployment

I argue that if linguistic equity is to be taken seriously, transparency must be a key element of responsible AI deployment. Users are better equipped to question and rectify detrimental outputs when they are aware of the model's limitations and biases. Communities most compacted by linguistic marginalization can take part in influencing the system's behavior due to this openness. In my opinion, this sort of transparency is crucial for fostering trust as well as preventing AI systems from maintaining the same linguistic hierarchies that are present outside of the digital realm.

Overall, my analysis of ChatGPT shows that developing inclusive and culturally grounded language models necessitates an ethical dedication to linguistic justice in addition to technical mastery. I contend that AI systems can start to challenge rather than reinforce current hierarchies of language and power by diversifying training data, reshaping reward structures to value non-standard linguistic forms, and guaranteeing transparent deployment. In the end, developing technologies that acknowledge, value, and elevate marginalized communities' linguistic identities, enabling their voices to fully and authentically exist in digital spaces.

CONCLUSION

Large language models like ChatGPT operate within linguistic, economic, and geopolitical systems that privilege dominant language,

standardized dialects, and Western cultural norms, leading to unequal benefits and harms across global linguistic communities. As this audit shows that the historical development of AI from Cold-War military research to corporate consolidation in Silicon Valley has produced structural conditions in which English and other high-resource languages become technologically legible, while thousands of marginalized, indigenous, and low-resource languages remain invisible. These inequities are produced through the data that trains the model, the labor practices that sustain it, and out puts it generates, often resulting in dialect erasure, unreliable performance for non-dominant languages, and standardized tone that flattens cultural variation. Such practices not only undermine linguistic diversity but also reinforce hierarchies tied to race, colonialism, class, and access to technological infrastructure.

Addressing these inequities requires reimagining AI systems as socially embedded technologies whose design decisions carry ethical and cultural consequences. Meaningful reform cannot be achieved merely through technical optimization or claims of neutrality; it must include community-centered data practices, dialect-preserving modeling strategies, and transparent accountability infrastructures that invite participation from marginalized linguistic groups. By valuing pluralism rather than imposing uniformity. AI systems can move beyond reproducing existing hierarchies and towards supporting linguistic justice, cultural representation, and equitable access to digital tools. Ultimately, the challenge is not only to build models that perform well across languages but to develop technologies that respect the identities, histories, and communicative practices of the communities they serve.

REFERENCES

- Bardol, F. (2025). ChatGPT Reads Your Tone and Responds Accordingly-Until It Does Not: Emotional Framing Induces Bias in LLM Outputs. *arXiv*, 2025, arXiv:2507.21083.
- Bellan, R. (2025). Sam Altman Says ChatGPT Has Hit 800M Weekly Active Users. *TechCrunch*, 6 Oct. 2025, www.techcrunch.com/2025/10/06/sam-altman-says-chatgpt-has-hit-800m-weekly-active-users.
- 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*, pp. 610–623. <https://doi.org/10.1145/3442188.3445922>
- Birhane, A. (2020). Algorithmic Colonialism. *SCRIPTed*, 17 (2), 389–409.
- Buscemi, A., Lothritz, C., Morales, S., Gomez-Vazquez, M., Claris'o, R., Cabot, J., & Castignani, G. (2025). Mind the language gap: Automated and augmented evaluation of bias in LLMs for high- and low-resource languages. *ArXiv*, abs/2504.18560.
- Cao, Yang Trista, Sotnikova, A., Zhao, J., Zou, L.X., Rudinger, & Daumé III, H.(2025). Multilingual large language models leak human stereotypes across language boundaries. *Proceedings of the Fourth Workshop on NLP for Positive Impact (NLP4PI)*, 2025.
- ChatGPT, the Generative AI Chatbot, Is Released. *History.com*, A&E Television Networks, www.history.com/this-day-in-history/chatgpt-released-openai. Accessed 5 Dec. 2025.
- Couldry, N., & Mejias, U.A. (2019). *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford UP.
- Dentella, V., Günther, F., & Leivada, E. (2023). *Systematic testing of three Language Models reveals low language accuracy, absence of response stability, and a yes-response bias*. *Proceeding of National Academic Science*, 120 (51) e2309583120, <https://doi.org/10.1073/pnas.2309583120>
- Dobariya, O., & Kumar, A. (2025). Mind Your Tone: Investigating How Prompt Politeness Affects LLM Accuracy. *arXiv*, 2025, arXiv:2510.04950.
- Fleisig, E., Smith, G., Bossi, M., Rustagi, I., Yin, X., & Klein, D. (2024). Linguistic bias in ChatGPT: Language models reinforce dialect discrimination. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 2024, pp. 13541–13564.
- Gray, M.L., & Suri, S. (2019). *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Houghton Mifflin Harcourt.
- Hinov, N. (2025). The energy hunger of AI: Large language models as challenges and enablers for sustainable energy. *Energies*,18(17), article 4701, www.mdpi.com/1996-1073/18/17/4701.
- Introducing OpenAI. *OpenAI*, 11 Dec. 2015, www.openai.com/blog/introducing-openai.
- Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M.(2020).The state and fate of linguistic diversity and inclusion in the NLP World. *Proceedings of the Association for Computational Linguistics*, 6282-6293.
- Longo, B. (2021). *Words and Power: Computers, Language, and U.S. Cold War Values*.Springer, 2021.
- Morrison, J., Na, C., Fernandez, J., Dettmers, T., Strubell, E., & Dodge, J. (2025). Holistically evaluating the environmental impact of creating language models. *arXiv*, 3 Mar. 2025, arxiv.org/abs/2503.05804.
- “OpenAI Claims to Have Hit \$10B in Annual Recurring Revenue.” *TechCrunch*, 9 June 2025, www.techcrunch.com/2025/06/09/openai-claims-to-have-hit-10b-in-annual-recurring-revenue.
- OpenAI*. Khosla Ventures, www.khoslaventures.com/portfolio/openai.
- Pakray, P., Gelbukh, A., & Bandyopadhyay, S. (2025). Natural language processing applications for low-resource languages. *Natural Language Engineering*,31(2),183-197.

- Perrigo, B. (2023). Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic. *Time*, 18 Jan. 2023, time.com.
- Sallam, M., Al-Mahzoum, K., Alshuaib, O., Alhajri, H., Alotaibi, F., Alkhurainej, D., Al-Balwah, M.Y., Barakat, M., & Egger, J.(2023). Language discrepancies in the performance of Generative Artificial Intelligence Models: An examination of infectious disease queries in English and Arabic. *BMC Infectious Diseases*, 23, 1–13.
- Sherma, A. (2023). Process of perfection method and its adaptability in EFL classrooms of Nepal. *SSRN*, 2023, ssrn.com/abstract=5331619.
- Singh, S. (2025). ChatGPT Users Stats (December 2025)-Growth & Usage Data. *DemandSage*, 20 Nov. 2025, www.demandsage.com/chatgpt-users-stats-growth-usage-data.
- Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 3645–3650.
- Thapa, M.B., & Shrestha, K. (2026). Empowering teachers through ChatGPT: Perceptions and experiences in English as a Foreign or Second Language (EFL/ESL) Classrooms as a Form of Community Service. *ASEAN Journal of Community Service and Education*, 5(1), 21–32.
- Thapa, M.B. (2025). Defying the Status Quo: Social and cultural struggles in James Baldwin’s *Sonny’s Blues*. *International Journal of Sciences and Innovation Engineering*, 2(6), 732–736.
- Weissweiler, L., Hofmann, V., Kantharuban, A., Cai, A., Dutt, R., Hengle, A., Kabra, A., Kulkarni, A., Vijayakumar, A., Yu, Schuetze, H., Oflazer, K. & Mortensen, D. (2023). Counting the Bugs in ChatGPT’s Wugs: A multilingual investigation into the morphological capabilities of a large language model. *arXiv*, 2023, arxiv.org/abs/2305.08005.
- Yin, Z., Wang, H., Horio, K., Kawahara, D., & Sekine, S. (2024). Should We Respect LLMs? A Cross-Lingual Study on the Influence of Prompt Politeness on LLM Performance. *arXiv*, 2024, arXiv:2402.14531.
- Zhu, S., Wang, W., & Liu, Y. (2024). *Quite good, but not enough: Nationality bias in large language models - a Case study of ChatGPT*. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 13489–13502.