

Writing Spaces

Assignments & Activities Archive

Gen AI: Large Language Models and Linguistic Whitewashing

Chris Mayer

This activity is a selection from the *Writing Spaces* Assignment and Activity Archive, an open access companion resource to the Writing Spaces open textbook series.

To access additional selections, visit: <http://writingspaces.org/aaa>.

Assignment and activity selections © 2024 by the respective authors. Unless otherwise stated, these works are licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0) and are subject to the Writing Spaces Terms of Use. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>, email info@creativecommons.org, or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA. To view the Writing Spaces Terms of Use, visit <http://writingspaces.org/terms-of-use>.

All rights reserved. For permission to reprint, please contact the author(s) of the individual articles, who are the respective copyright owners.

Gen AI: Large Language Models and Linguistic Whitewashing

Chris Mayer

Overview

This multi-step lesson engages students in the critical analysis of Generative Artificial Intelligence (Gen AI) and its impact on linguistic diversity. Through assigned readings, writing tasks, and in-class activities, students will explore how Large Language Models (LLMs) can reinforce linguistic hierarchies and biases.

Even older LLMs, like GPT-3, can generate content that human judges find comparable or superior to human writing in terms of fluency, coherence, relatedness, logic, and interest (Xie et al.), although they struggle to capture and express deep emotional understanding (Petroşanu et al.). LLMs have quickly become integral to education, and Gen AI is now part of the way students write in university (Dobrin 2023). It also holds the potential to transform peer review and feedback in writing-intensive courses (Mayer 2024).

On one hand, tools like ChatGPT have the potential to democratize access to linguistic resources and guide marginalized students in navigating Standard Academic English (SAE). However, even as they do so, they perpetuate and entrench SAE norms, further marginalizing non-standard dialects and failing to account for the socio-ethnolinguistic positionality of users. This can have powerful effects on writers, as we deeply embody our language (e.g., Baker-Bell 2020).

This three-part mini-unit aims to familiarize students with LLMs and their relationship to linguistic diversity. First, students will engage

with background readings on LLMs and linguistic diversity, followed by reflective writing. Next, for homework, they will [watch a video lecture](#) titled "Gen AI LLMs Washing Out Linguistic Features of Appalachian and Black Englishes" available on YouTube, or a similar instructor-recorded lecture. Finally, in class, students will experiment with LLMs using non-standard English dialects and discuss the social justice implications of their findings.

Time Commitment

About a third of a class period (25 minutes) and a full class period (75 minutes).

Materials

Computers or tablets with internet access

Access to Google Drive and several Large Language Models like ChatGPT, Bard, Gemini, and Copilot

Activity Process

Part I – Homework and half a class (25 minutes):

- Assign one or more background readings which discuss LLMs and/or relate to linguistic diversity and justice such as the ones below, which are arranged in order from easy to hard:
 - *What ChatGPT Means for Linguistic Diversity and Language Learning* (Sharma), selections (especially Chapter 3) from *Linguistic Justice: Black Language, Literacy, Identity, and Pedagogy* (Baker-Bell, 2020), *Navigating the New Frontier of Generative AI in Peer Review and Academic Writing* (Mayer) and *The Power Chapter* (D'Ignazio and Klein).
- Students will write a three part, three-paragraph response to

at least one assigned reading. Sample prompt:

- “Write three paragraphs and at least one page. Paragraph one should be a summary of the main argument from the reading; paragraph two should include personal reflections on how the reading relates to your own experiences with language; paragraph three should lay out your initial thoughts on the relationship between LLMs and linguistic diversity.”
- Finally, in class, students will get into small groups and share their personal reflections and their thoughts on the relationship between LLMs and linguistic diversity. Then, the instructor will ask groups to share the connections between LLMs and linguistic diversity with the whole class.

Part II – Homework Prep:

- As homework, students will watch the video lecture on [LLMs washing out linguistic features of Appalachian and Black Englishes](#) or something similar made by the instructor. The linked video above connects AI image generators to LLMs.
 - They will take notes and create a “highlights outline” summarizing key points of the lecture and also write a one paragraph reaction to the lecture, reflecting on how the content relates to the earlier reading and their understanding of language diversity. This can be hand written or turned in online.

Part III – Full class (75 minutes):

- Students will form groups of three to four and first share

their reactions to the video lecture, including the connections they wrote about.

- Then, groups should make a shared Google Doc. Each group will choose a nonstandard English dialect (e.g., AAVE, Appalachian English, Cajun English, Hawaiian Pidgin) and then search the internet for authentic written samples and, as a group, select three or four sample sentences. They should paste the sentences into their Google Doc.
- Next, each group member will choose one LLM to work with (Gemini, Bard, ChatGPT, Claude, etc.). They will input the nonstandard English sentences they chose and ask for changes or revisions, making sure to ask non-leading questions that do not ask for “corrections” or “improvements” but rather edits or revisions. For instance, they might ask “How would you revise this?” but not ask “How would you make this better?” Each student should make notes about the commentary their LLM provided in the Google Doc. Then, groups will analyze the responses, focusing on how each LLM handles the nonstandard English. They should list out the patterns and trends they found, with illustrative examples, and prepare a short presentation of their findings either in the Google Doc or with Slides.
- After ~30 minutes, two groups will volunteer/be selected to present their findings.
- Then, back in small groups, students will discuss the following questions:
 - 1. How do LLMs' tendencies to standardize nonstandard English dialects impact speakers of these dialects? 2. What are the potential consequences of linguistic homogenization in

educational and professional contexts? 3. How might the biases in AI language processing technologies perpetuate existing social and linguistic inequalities? 4. What ethical responsibilities do developers of AI language technologies have in preserving linguistic diversity? 5. How can AI be designed or regulated to better respect and reflect linguistic diversity? 6. What are some benefits and drawbacks of having a “broader audience” as defined by Gen AI’s standardization efforts?

- After 10 or 15 minutes, call the class back together and ask each group for their main takeaway point from this lesson. Give groups 2 minutes to talk out/prepare what their point is, and then ask each group to share.
- Finally, assign a writing prompt as homework that encourages reflection and thought about LLMs and linguistic equity, such as this: “In a future increasingly dominated by AI, what steps can we take to ensure that technological advancements contribute to, rather than detract from, linguistic and social equity?”

Learning Outcomes

Students engaging in this activity/assignment will:

- Develop an understanding of the impact of LLMs on linguistic diversity.
- Critically analyze how LLMs interact with nonstandard varieties of English.
- Reflect on the social justice implications of language standardization by AI technologies.

Learning Accommodations

- Choose a background reading appropriate to the level of the class or ask students to browse three reading options and choose the one they are most comfortable with and interested in reading.
- For internet readings and non-PDF readings that can be copied, remind and/or show students that MS Word has a “read aloud” function under the Review tab, and that they can copy and paste the reading text and listen if they process better through listening.
- For the video lecture on YouTube, remind and/or show students about the closed captioning option.
- Consider previewing concepts such as dialects and ethnolects, systematic grammar rules, and embodiment of language.

Works Cited

- Baker-Bell, April. *Linguistic Justice: Black Language, Literacy, Identity, and Pedagogy*. Routledge, 2020.
- D'Ignazio, Catherine, and Lauren Klein. "The Power Chapter." *Data Feminism*, MIT Press, 2020. <https://data-feminism.mitpress.mit.edu/pub/vi8obxh7>
- Dobrin, Sidney I. *Talking About Generative AI: A Guide for Educators*. Version 1.0, Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, 2023, <https://files.broadviewpress.com/sites/uploads/sites/173/2023/05/Talking-about-Generative-AI-Sidney-I-Dobrin-Version-1.0.pdf>
- Mayer, Chris. "Navigating the New Frontier of Generative AI in Peer Review and Academic Writing." *Teaching and*

Generative AI: Pedagogical Possibilities and Productive Tensions, edited by Beth Buyserie and Travis N. Thurston, Utah State University, 2024,
<https://uen.pressbooks.pub/teachingandgenerativeai/chapter/navigating-the-new-frontier-of-generative-ai-in-peer-review-and-academic-writing/>

Petroșanu, Dana-Mihaela, Alexandru Pîrjan, and Alexandru Tăbușcă. "Tracing the Influence of Large Language Models across the Most Impactful Scientific Works." *Electronics*, vol. 12, no. 24, 2023, article 4957,
<https://doi.org/10.3390/electronics12244957>

Sharma, Yojana. "What ChatGPT Means for Linguistic Diversity and Language Learning." *University World News*, 24 Feb. 2023,
<https://www.universityworldnews.com/post.php?story=20230223132734294>

Xie, Zhouhan, Trevor Cohn, and Jey Han Lau. "The Next Chapter: A Study of Large Language Models in Storytelling." *Proceedings of the 16th International Natural Language Generation Conference*, edited by C. Maria Keet, Hung-Yi Lee, and Sina Zarrieß, Association for Computational Linguistics, 2023, pp. 323-51,
<https://aclanthology.org/2023.inlg-main.23>

Further Reading

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Polity, 2019.

Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World*. MIT Press, 2018.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. NYU Press, 2018.