

An Exploration of Generative Artificial Intelligence’s Linguistic Capabilities: A Case Study in African American English

Tania Mishra – Computer Science, Cognitive Science
Faculty Advisor: Dr. Michael Zimmer

INTRODUCTION

The United States has a long history of using language as a tool of oppression. There are many different variations of English: **Standard American English (SAE)**, Standard British English, **African American English (AAE)**, Spanglish, and more. These variations can be treated as a separate language; each has specific vocabulary and grammar rules. SAE is the preferred language in the U.S., which marginalizes those who speak differently, especially speakers of AAE. Generative Artificial Intelligence (GAI) models are trained on large datasets of human-generated texts collected from the internet, resulting in **training sets predominantly made up of SAE text**. This study aims to explore popular GAI models’ ability to use AAE.

RESEARCH QUESTIONS

- How do GAI models **translate SAE to AAE** when requested to **rewrite** a response in AAE?
- How well do GAI models **use AAE** grammar rules and vocabulary when requested to **respond** in AAE?

GAI MODELS

- ChatGPT by OpenAI
- Claude by Anthropic
- Gemini by Google
- Copilot by Microsoft
- Aya by Cohere
- Llama by Meta

PROMPT DEVELOPMENT

- 30 prompts
- Loosely grouped into 6 categories (*assigned during data analysis*)
 - Black Culture, Idea Generation, Personal, Professional, General Information, Creative

METHODOLOGY

For each GAI model:

- Created a dummy account so that model did not have prior search history
- Gathered the “Control” responses: prompted model with the prompts, recorded responses
- Gathered the “Rewritten” responses: prompted model to rewrite the control response in AAE in the same chat as the control
 - “Please rewrite that in African American English.”
- Cleared all chat history
- Gathered the “Respond” responses: prompted model to respond in AAE
 - “Please respond in African American English.” added to the end of the prompt

REFERENCES

- Blodgett, S. L., Green, L., & O'Connor, B. (2016). Demographic Dialectal Variation in Social Media: A Case Study of African-American English. Association for Computational Linguistics, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 1119–1130. <https://doi.org/10.18653/v1/D16-1120>
- Smitherman, G. (1999). Talkin that Talk: Language, Culture and Education in African America. Routledge.
- Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2018). Dialect prejudice predicts AI decisions about people's character, employability, and criminality. ArXiv. <https://doi.org/10.48550/arXiv.2403.00742>

RESULTS

Example Responses (*Copilot*)

Q: “Write me a job interview thank you note.”

“Control”

Subject: Thank You for the Interview
Dear [Interviewer's Name],
Thank you for taking the time to meet with me today. I appreciate the opportunity to discuss the [position name] role at [company name]. Our conversation reinforced my enthusiasm for joining your team and contributing to [specific project or goal discussed]. I look forward to the next steps in the process and hope to hear from you soon. Please feel free to reach out if you need any additional information.
Thank you again for considering my application.
Sincerely,
[Your Full Name]

Percent Changed:
70.68%

“Rewritten”

Subject: Thank You for the Interview
Yo, [Interviewer's Name],
I just wanna give a shout-out for the time you spent with me today. I appreciate the chance to chop it up about that [position name] gig at [company name]. Our convo got me even more hyped about joining your crew and making moves on that [specific project or goal discussed].
I'm lookin' forward to the next steps in the process, and I hope you'll hit me up soon. If you need any more info, holla atcha girl.
Thanks again for checkin' out my application.
Stay blessed,
[Your Full Name]

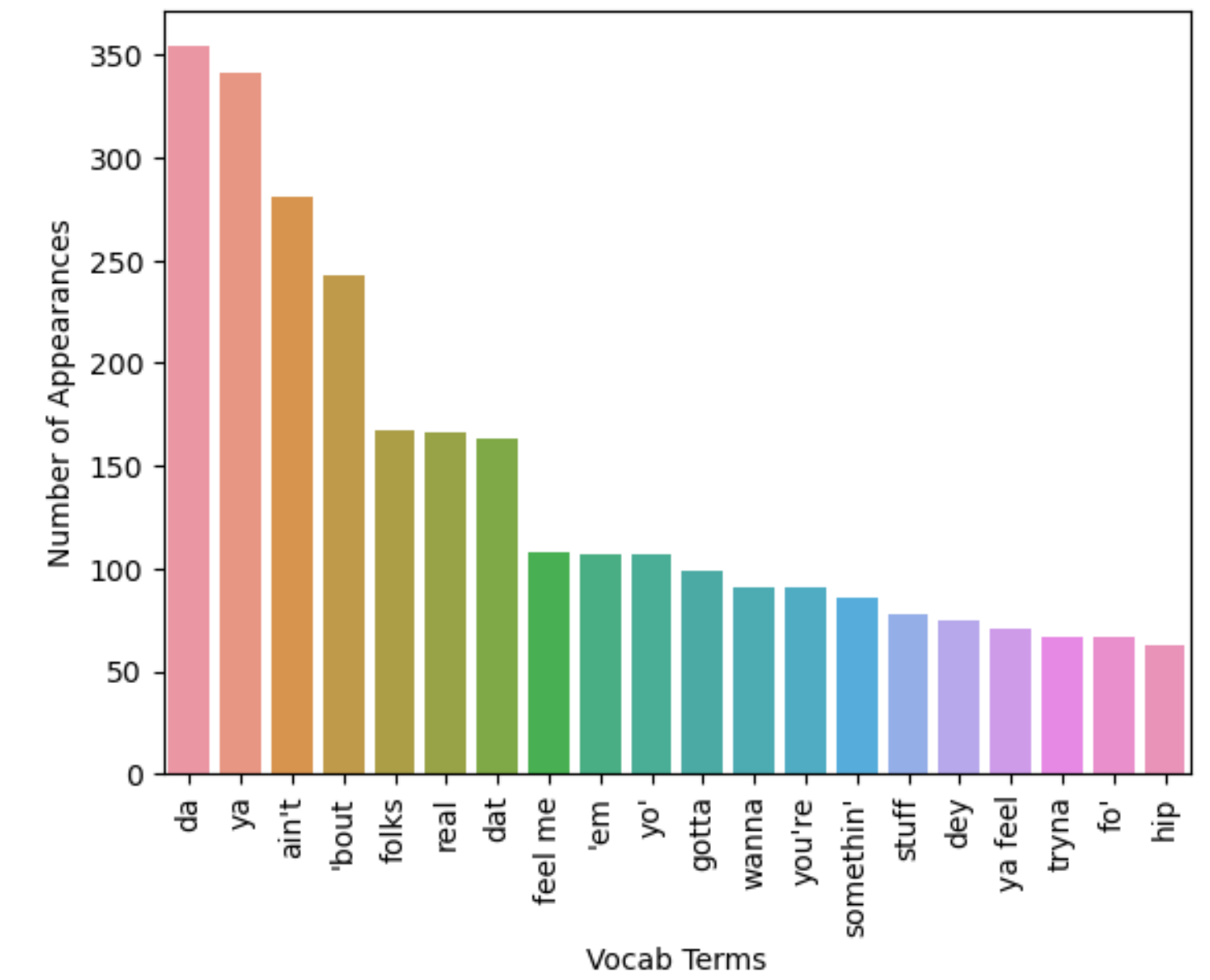
AAE Score:
15.63%

“Respond”

Subject: Thank You for the Interview Opportunity
Hey [Interviewer's Name],
I wanted to take a moment to express my gratitude for the opportunity to interview with [Company Name]. It was great connecting with you and learning more about the team and the role. I appreciate your time and thoughtful questions.
I'm excited about the possibility of joining [Company Name] and contributing to the team's success. Please feel free to reach out if you need any additional information from me.
Thanks again, and I look forward to hearing from you soon!
Best regards,
[Your Full Name]

AAE Score:
3.74%

Frequency of AAE Vocabulary (among the “Rewritten” and “Respond” responses):



Similarity and Differences (*between the “Control” and “Rewritten” responses*):

- Computed the lexical similarity (and difference) using the Jaccard Index

	Personal	Professional	Black Culture	General Information	Idea Generation	Creative	Average
ChatGPT					35.55%		54.63%
Claude						83.03%	80.29%
Gemini							70.12%
Cohere							61.70%
Copilot							63.02%
Llama							70.41%
Average	61.98%	69.76%	75.11%	72.97%	59.59%	61.87%	66.70%

	Color
Percent Changed	
0-20%	
20-40%	
40-60%	
60-80%	
80-100%	

AAE Scores (*among the “Rewritten” and “Respond” responses*):

- Computed the score using the model from The Sociostatistical Language Analysis Lab
 - Model is trained on user-generated (African American demographic) data from Twitter
 - Checks for lexical, phonological, and syntactic/grammatical variations specific to AAE

	Personal	Professional	Black Culture	General Information	Idea Generation	Creative	Average
ChatGPT							16.27%
Claude							26.37%
Gemini							17.46%
Cohere			39.21%				24.39%
Copilot							15.73%
Llama		7.30%					22.54%
Average	20.35%	11.21%	28.02%	24.91%	11.78%	26.49%	20.46%

	Color
AAE Percentage	
0-20%	
20-40%	
40-60%	
60-80%	
80-100%	

CONCLUSION

This exploration revealed that these GAI models often do not use AAE accurately, indicating bias towards the African American community. Further prompt development, including more prompts and multiple trials, could improve the robustness of this study. Future work could also include looking at how AAE speakers interact with GAI models.