

Chatbot Architecture

Artificial intelligence chatbots and virtual reality assistants rely heavily on large language models to generate relevant and appropriate responses based on user input. Notably, transformer architecture creates correlations and associations between words, or token collocation,¹ to model and anticipate the subsequent words, ideas, and phrases in response to user input.² However, the generation model is trained through a corpus,³ or a large set of text data (primarily online). Therefore, newer adaptations of chatbots become more accurate through the correction of core algorithms and corpora expansion.

Argument: Corpus restrictions and constraints can generate artificial intelligence bias in chatbots.

Objectives

Investigating the architecture responsible for the limitations of artificial intelligence and chatbots by exploring:

- The relationship between *corpus* adjustments and bias production in artificially generated text.
- The creative limitations of *chatbots* for occupations requiring creative and inventive design.
- Solutions to expanding corpora for *low-recourse languages*.

Methods

LancsBox 6.0:

The collocation of tokens was graphed for the American and British English corpora, notably for the word "race."

People and A.I. Research Tools:

The B.E.R.T. tool predicted thirty potential characterizations of people from Los Angeles and New York demographics.

OpenAI's ChatGPT:

The conversational language capacity of the chatbot was tested in low-resource and indigenous languages, primarily Cherokee and K'iche.'

Specialist Interviews:

Professors and field experts were contacted to discuss the potential for chatbots to advance their fields in mathematics and law, respectively, following questionnaires.

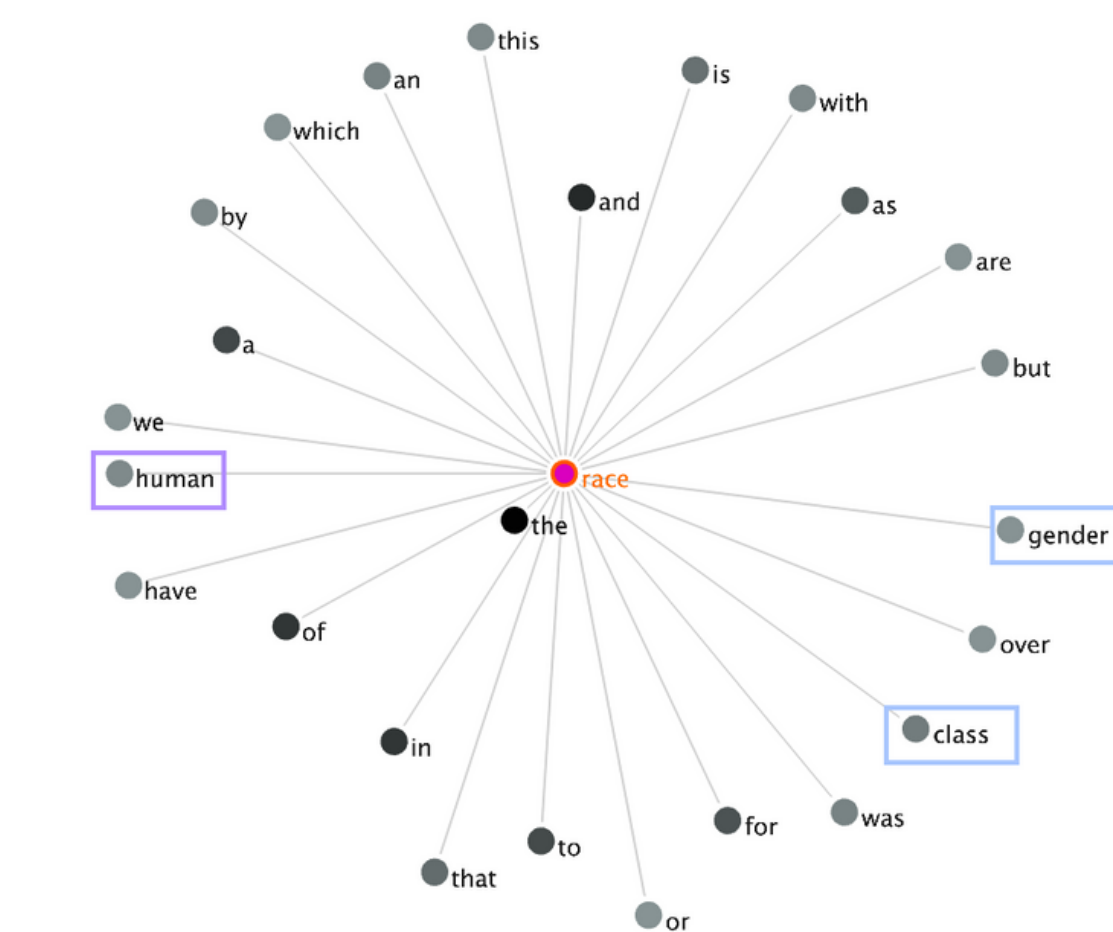
LancsBox Token Collocation

Token collocation characterizes the tendency for words, or tokens, to co-occur, indicating a specific connotative significance. Language models and tools, however, can locate collocations. For instance, the Lancaster University corpus toolbox demonstrates the collocates found within the American⁴ and British English corpora⁵ below. Specifically, when locating collocates for "race," American English encounters tokens more associated with societal design, while British English encounters tokens more associated with interpersonal communication.

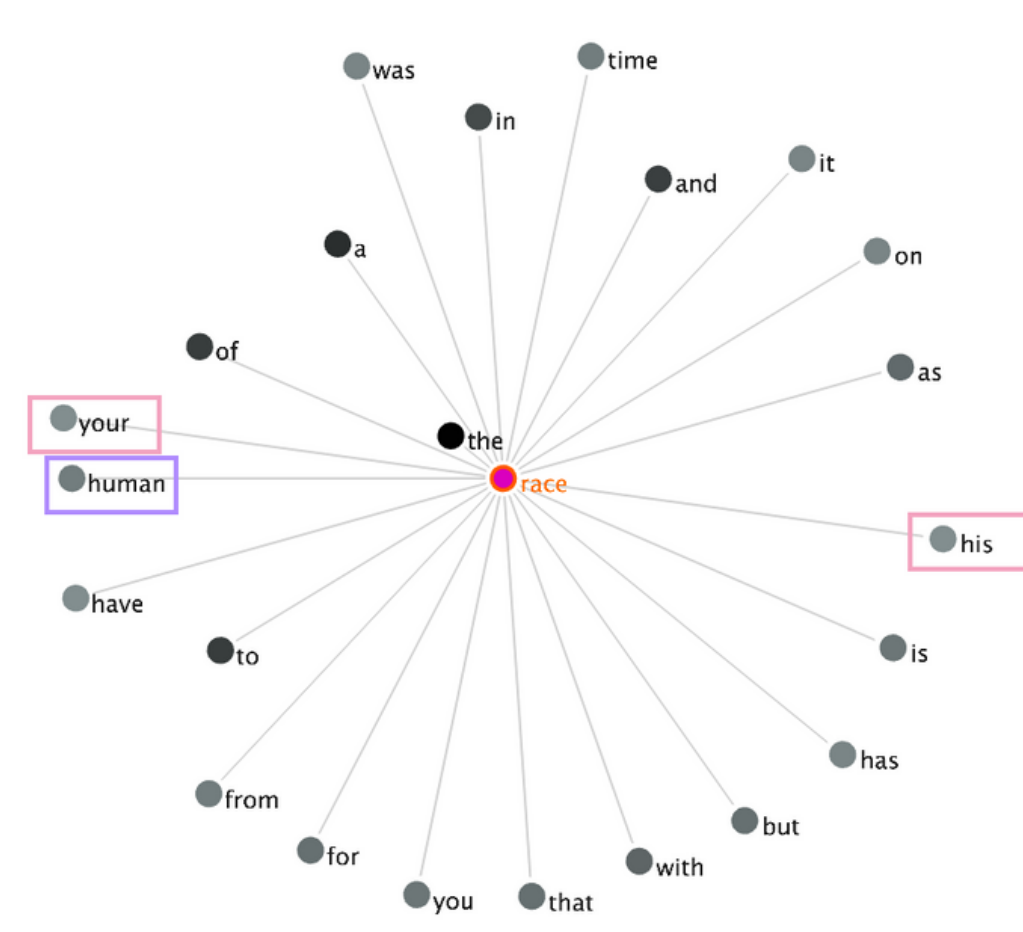
Nathan Maldonado, Emily Gebhardt, Astrid Giugni*

Department of Computer Science, Duke University, Durham, NC

American English Corpus (2006): "Race"
GraphCollocation

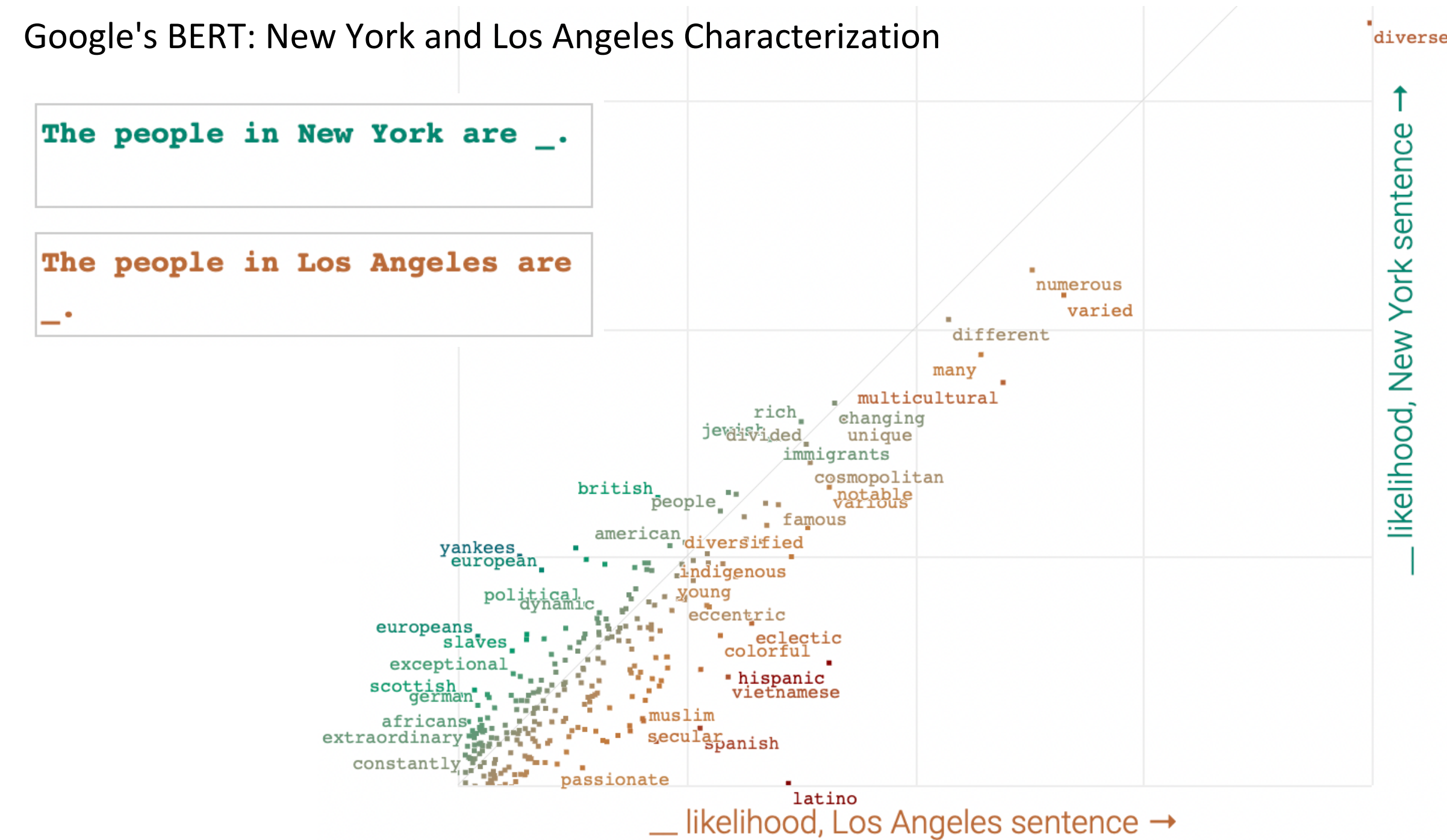


British English Corpus (2006): "Race"
GraphCollocation



B.E.R.T. Prediction Modeling

Bidirectional Encoder Representations from Transformers (BERT) considers bidirectional context to prepare appropriate responses. Specifically, the model is trained through various corpora after establishing relationships and correlations amongst tokens—collocations. While training, however, the model begins to mask words and predict these masked words to determine response sensibility through what is known as Masked Language Modeling and Next Sentence Predictions.



Above, Google's B.E.R.T model was tasked with filling in the masked word for the sentence: the people in [Los Angeles or New York] are [adjective] . The model generated the likelihood of each token filling the mask, indicating prediction biases amongst each city. For instance, people from Los Angeles were more likely to be characterized by "multicultural," "eccentric," and "young" adjectives. In contrast, people from New York were more likely to be characterized by "European," "yankee," and "wealthy" adjectives.

Creative Limitations

Brandon Garret, a glass-box advocate and Director of the Wilson Center for Science and Justice at Duke University, anticipates the utility of chatbot consultation. However, Dr. Garrett referenced the various boundaries to artificial intelligence bias infiltrating a fair court. For instance, judges may consult sources to reach conclusions and interpret the law but still rely on clerks and treatises for reviews.

"There is not necessarily a need for [a judge's] description of legal rules to be original.

- *Brandon Garrett, J.D.*

Additionally, Clark Bray, an Associate Professor of the Practice of Mathematics at Duke University, emphasized the relevance of ChatGPT but reconciled it with the novelty of the tool and its potential applications in mathematics.

Translation Limitations

For languages with extensive corpus data or low-level languages, ChatGPT and other chatbots have more to rely on to perform these tasks. Therefore, when asking ChatGPT to interpret Shakespeare's Sonnet 116, the chatbot could not respond in Cherokee or K'iche'. Instead, the chatbot alerts users of its limitations and demonstrates an understanding of the potential for bias production and inaccurate media representations.⁶

"Cherokee is a Native American language with its unique writing system, and accurate translation requires deep linguistic and cultural knowledge."

"I'm unable to provide an accurate translation of the text into K'iche' since it's a complex task that requires a deep understanding of the language and its nuances."

- ChatGPT (2023)

- ChatGPT (2023)

Nevertheless, the chatbot can produce text responses for other low-written resource languages, so while these warnings are helpful, they can be user dependent.

Conclusion

Corpus limitations through data manipulation or inaccessibility encourage bias infiltration in chatbot outputs. Future research areas include exploring the utility of chatbots in creative fields through assistant generative text. Lastly, speech-to-text synthesis technology, human interpreters, and archival collection must be utilized to expand corpora for low-resource languages and expand chatbots' utility.

Acknowledgements

The author would like to thank Duke University's Initiative for Science & Society. Specifically, through concentrated attention under the UNIV103: Let's Talk About Digital You course, the project was strengthened by mentorship, scholarship, and fellowship. Additionally, the author would like to acknowledge the Huang Fellows Program for project guidance, specifically program directors Michael B. Waitzkin, J.D., LL.M., and Ron Grunwald, Ph.D.

References

1. Skansi, S. (Ed.). (2020) 2. People and AI Research. (2019) 3. Desagulier, G. (2014) in J. Robinson & D. Glynn (Eds.) 4. Potts, A., & Baker, P. (2012) 5. Baker, P. (2009) 6. OpenAI. (2023)