

The Promise and Peril of Large Language Models for Cultural Analytics

David Bamman

Associate Professor, School of Information, University of California, Berkeley

Abstract

Much work at the intersection of NLP and cultural analytics/computational social science is focused on creating new algorithmic measuring devices for constructs we see encoded in text (including agency, respect, and power, to name a few). How does the paradigm shift of large language models change this? In this talk, I'll discuss the role of LLMs (such as ChatGPT, GPT-4 and open alternatives) for research in cultural analytics, both raising issues about the use of closed models for scholarly inquiry and charting the opportunity that such models present. The rise of large pre-trained language models has the potential to radically transform the space of cultural analytics by both reducing the need for large-scale training data for new tasks and lowering the technical barrier to entry, but need care in establishing the reliability of results.

Bio


David Bamman is an associate professor in the School of Information at UC Berkeley, where he works in the areas of natural language processing and cultural analytics, applying NLP and machine learning to empirical questions in the humanities and social sciences. His research focuses on improving the performance of NLP for underserved domains like literature (including LitBank and BookNLP) and exploring the affordances of empirical methods for the study of literature and culture. Before Berkeley, he received his PhD in the School of Computer Science at Carnegie Mellon University and was a senior researcher at the Perseus Project of Tufts University. Bamman's work is supported by the National Endowment for the Humanities, National Science Foundation, an Amazon Research Award, and an NSF CAREER award.

CLiC-it 2023: 9th Italian Conference on Computational Linguistics, Nov 30 – Dec 02, 2023, Venice, Italy

✉ dbamman@berkeley.edu (D. Bamman)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)