

Application of Topic Modeling on Artificial Intelligence Studies as a Foundation to Develop Ethical Guidelines in African American Dementia Caregiving

Sunmoo YOON^{a,1}, Peter BROADWELL^b, Frederick F SUN^a,
Maria DE PLANELL-SAGUER^a and Nicole DAVIS^c

^aGeneral Medicine, Department of Medicine, Columbia University, New York, NY, USA

^bCenter for Interdisciplinary Digital Research, Stanford University, Stanford, CA, USA

^cSchool of Nursing, Clemson University, Clemson, SC, USA

Abstract. We applied natural language processing and topic modeling to publicly available abstracts and titles of 263 papers in the scientific literature mentioning AI and demographics (corpus 1 before Covid-19, corpus 2 after Covid-19) extracted from the MEDLINE database. We found exponential growth of AI studies mentioning demographics since the pandemic (Before Covid-19: N= 40 vs. After Covid-19: N= 223) [forecast model equation: $\ln(\text{Number of Records}) = 250.543 * \ln(\text{Year}) + -1904.38$, $p = 0.0005229$]. Topics related to diagnostic imaging, quality of life, Covid, psychology, and smartphone increased during the pandemic, while cancer-related topics decreased. The application of topic modeling to the scientific literature on AI and demographics provides a foundation for the next steps regarding developing guidelines for the ethical use of AI for African American dementia caregivers.

Keywords. Artificial intelligence, dementia caregiving, health disparity, ethics

1. Introduction

African Americans experience disproportionately higher rates of dementia than other racial groups [1]. We recently developed a lexicon-based African American Tweet detection algorithm using AI based on 3,291,101 Tweets and 14,692 works of literature to inform of culturally sensitive Twitter-based social support interventions for African American family dementia caregivers [2]. Through this algorithm, we intended to identify existing African American communities and discourse patterns on social media platforms like Twitter to understand community and culture (e.g., culturally sensitive expressions, norms). Such an algorithm would also enable us to reach the target population without wasting public resources (e.g., decreased screening burden and errors, increased safety).

¹ Corresponding Author: Sunmoo Yoon, PhD, MS, RN, 630 W 168th Street, PH105, New York, NY, 10032, USA; E-mail: sy2102@cumc.columbia.edu.

Considering the potentially unresolved distrust towards scientists and the U.S. government among African Americans; and their vulnerabilities to the widening digital divide since Covid-19, guidelines for the ethical use of artificial intelligence (AI) are urgently needed [2]. Therefore, this study aimed to apply natural language processing and topic modeling to understand trends and topics in AI and demographics before and during the Covid-19 pandemic to inform guideline development for the ethical use of AI in the social media domain. This study will provide a comprehensive summary of state of the art in scientific research as a foundation to develop guidelines for the ethical use of artificial intelligence, to be observed when implementing algorithms for social media-based interventions and recruitment regarding dementia caregiving.

2. Methods

We applied natural language processing (NLP) and topic modeling on a publicly available English-language scientific literature corpus of papers mentioning demographics and AI [3].

First, we collected and extracted the publicly available text of the abstracts, keywords and titles of the 263 articles relevant to demographics and AI (keywords: demographic and AI in abstract or MeSH indexing) from the MEDLINE database (corpus 1: prior to the Covid-19 pandemic – before February 28, 2020, corpus 2: during/after Covid-19 – March 1, 2020-March 27, 2023).

We built a forecast model to predict the number of studies in the field of demographics and the application of AI using a linear trend model computed from a natural log of the sum of the number of records (actual and forecast), given the natural log of the publication time.

Second, we applied natural language processing (NLP) to the texts using AutoMap and Perl expressions with the following steps: a) clean (e.g., remove extra spaces, convert British to American spelling, remove symbols, remove pronoun), b) preprocess (e.g., remove single letters, prepositions, noise verbs, numbers, and stop words, apply stemming and n-gram conversion), and c) refine (remove punctuation, lower case conversion, concept tuning and merging by a domain expert) [3].

Third, we applied the Newman clustering algorithm to two publicly available literature corpora followed by visualizing topics in a network diagram (corpus one before Covid-19 Newman modularity: 0.732, corpus two after Covid-19 Newman modularity: 0.639). Last, domain experts in African American culture studies, dementia caregiving and behavioral science performed interpretation on the detected topics. The parameters of this study were to analyze publicly available text from a public library, that does not meet the definition of human subject research requiring IRB review. Resources and analytic codes are available on GitHub and OSF.io (<https://osf.io/qruf3>).

3. Results

We found exponential growth of AI studies mentioning demographics since the pandemic (Before Covid-19: N= 40 vs. After Covid-19: N= 223) in the MEDLINE database [forecast model equation: $\ln(\text{Number of Records}) = 250.543 \cdot \ln(\text{Year}) + -1904.38$, $p = 0.0005229$] (Figure 1).

Eleven topics extracted from corpus one before Covid-19 (clustering coefficient: 0.732) include 1) patient data and clinical database, 2) aging and middle-aged males and females, 3) critical care and medication adherence, 4) cancer biomarker and drug therapy, 5) disease risk assessment and telemedicine, 6) physiopathology and diagnostic imaging, 7) chronic disease survival rate and mortality, 8) diagnosis decision support and epidemiology, 9) pharmacokinetics, 10) adverse effect and smoking and drinking behavior, 11) global health and legislation. Six topics detected from corpus two after Covid-19 (clustering coefficient: 0.639) include 1) treatment outcome and smartphone, 2) deep learning/neural network and disease risk factors and prognosis, 3) aging, quality of life, and psychology (anxiety and depression), 4) genetics and Covid diagnosis, 5) adverse effect and radiography, 6) blood biomarker. The top five emerging concepts (dots in Figure 2) include diagnostic imaging (relative frequency: 0.095 \rightarrow 0.219), quality of life (0.071 \rightarrow 0.191), Covid (0 \rightarrow 0.098), psychology (0.023 \rightarrow 0.049), smartphone (0 \rightarrow 0.018).

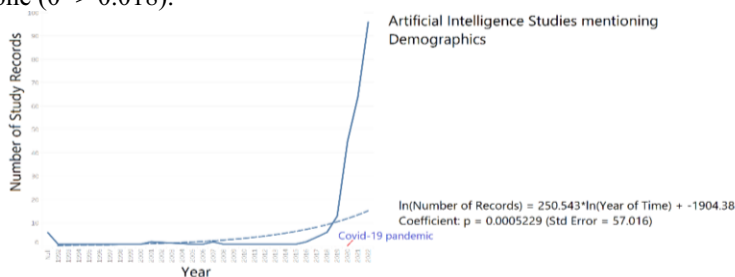


Figure 1. The exponential growth of AI studies mentioning demographics since the pandemic (Before Covid-19: N= 40, After Covid-19: N= 223) in the MEDLINE database and forecast model

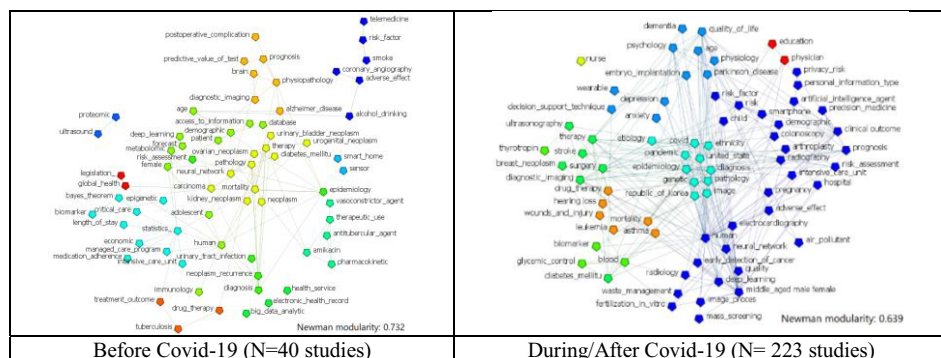


Figure 2. Topics extracted from two corpora of scientific literature on AI mentioning demographics before (Left) and during/after Covid-19 (Right)

4. Discussion and Conclusion

This study explored the trends and topics in the scientific literature related to AI and demographic in an effort to inform the development of ethical guidelines for the application of AI among African Americans. One topic with a notable increase during the Covid-19 was robot use in nursing care. The Covid-19 pandemic has impacted nurse staffing, especially for dementia patients within nursing homes [2]. Technological advancement may explain the increase of AI use as a new means of interaction

worldwide [4]. The interaction with robots that once was seen as science fiction is now part of a new reality, the new normal. Although AI may provide solutions for the increased demands of the growing elderly population and the nursing shortage, AI-based tools have proliferated without fully addressing ethical dilemmas [4]. Careful measures of potential benefits and risks must occur before adopting such technologies [2]. As our topic modeling detected the growing interest in genetic profiling for chronic diseases, gene profiling presents ethical, legal, and social challenges in privacy and confidentiality, accuracy and test implementation and implications for African American family caregivers of a person with dementia [5]. Despite in the U.S. the Genetic Information Nondiscrimination Act (GINA)² [Public Law 110-233] and other laws prohibit genetic and other discrimination, these laws have significant limitations, and some provisions are under threat, such as stigma and discrimination in the context of health insurance or employment for African American dementia caregivers [2, 4].

Another topic that emerged during the pandemic was AI use in mental health [4]. Suicide and anxiety are debilitating mental health issues that affect many African American dementia caregivers [1]. As the use of AI in caregiving continues to grow, it is crucial to consider the ethical implications and potential unintended consequences for this vulnerable population [2,4]. For example, AI should be designed with cultural sensitivity and awareness of racial disparities in mind and should aim to enhance rather than replace human connection and support for dementia caregivers [2,4]. The exponential growth of AI applications in healthcare and caregiving during COVID-19 has the potential to transform the way African American dementia caregivers provide care [4]. Ultimately, AI applications in caregiving should prioritize the well-being and dignity of African American dementia caregivers and their loved ones [2].

In conclusion, we found that AI concerning demographics has exponentially grown in the health domain since Covid-19. While topic modeling in this study detected the terms diagnostic imaging, quality of life, Covid, psychology, and smartphone as emerging topics from the scientific literature of AI studies mentioning demographics, there is a dearth of evidence on the ethical use of specific AI technologies, and demographic detection algorithms in particular, in dementia caregiving research. Ethical guidelines for using such technologies based on rigorous evidence are urgently needed to reduce health disparities in dementia caregiving.

Acknowledgments

U.S. federal grants R01AG060929 and R01AG060929-04S1 (PI: Yoon)

References

- [1] Alzheimer's Association. Race, ethnicity and Alzheimer's in America, 2021.
- [2] Broadwell P, Davis N, Yoon S. Using AI to develop a lexicon-based african american tweet detection algorithm to inform culturally sensitive twitter-based social support interventions for african american dementia caregivers. *Stud Health Technol Inform*. 2022 Jan;289:1-4.
- [3] Chowdhary K, Chowdhary KR. *Natural language processing: fundamentals of artificial intelligence*. Jodhpur: Springer Nature Press; 2020. 603 p.
- [4] Jobin A, et al. The global landscape of AI ethics guidelines. *Nat. Mach. Intell*. 2019 Sep;1(9):389-99.
- [5] Chen IY, Joshi S, Ghassemi M. Treating health disparities with artificial intelligence. *Nat. Met*. 2020 Jan;26(1):16-7.