"Would You Rather Have It Be Accurate or Diverse?" How Male Middle-School Students Make Sense of Algorithm Bias

Golnaz Arastoopour Irgens, Clemson University, garasto@clemson.edu
JaCoya Thompson, Northwestern University, jacoyathompson2020@u.northwestern.edu

Abstract: As data-driven decisions become more ubiquitous, it will be critical for youth to understand the impacts of algorithm bias. In this study, we discuss the design of an extra-curricular data science program and examined how the participants (12 males, ages 11 - 13) made sense of algorithm bias and discrimination. We conducted a critical discourse analysis on one classroom discussion. Results suggest that participants showed initial understandings that algorithms contain biases that may perpetuate discrimination.

Introduction and theory

Big data algorithms are becoming increasingly relevant in social and policy decisions and can now be used to decide whether people secure employment, receive loans, or are convicted of a crime. Unfortunately, such algorithms reinforce racial and gender discrimination at large scales and are largely unintelligible to the public (Noble, 2018; O'Neil, 2016). As these big data-driven decisions become more ubiquitous, it will be critical for young people to understand computational statistics and the ethical consequences of algorithm bias. Recent studies have examined big data practices for middle-school students (Jiang & Kahn, 2019) and the intersections of data visualizations and spatial justice (Rubel et al., 2017). However, few have investigated how young people engage discursively with big data algorithm bias and racial/gender discrimination. In this study, we designed a data science extra-curricular course for middle school students. The research question is *How do male middle-school students engage in algorithm bias and discrimination topics in a data science curriculum?*

Philip and colleagues (2013) have proposed a framework on big data for democratic participation that is rooted in both sociocultural and critical pedagogy theories. Using these two perspectives, Philip and colleagues proposed three categories of student objectives for democratic data science: content proficiency and discursive fluency (the ability to use language and tools of the discipline), motivated use of content (learners see themselves as users of data science and work towards greater justice and equity in society), and the politics of knowledge (learners know that data is political and address limitations/opportunities for particular populations). These three categories purposefully center topics of inequities, power, and ethics around big data algorithms.

Methods

The participants in this research were 12 North-American males, ages 11-13, who enrolled in the course. Five students were European-American, four were Indian-American, two were Asian-American, and one was Latino. All students self-reported that they had previous experience with programming. Audio recordings were collected of student participation in three discussion-based activities, and all names were changed to pseudonyms. We were the instructors of the course and former high school math and computer science teachers. The course took place on six consecutive Saturdays for 2.5 hours each day. The first half was an introduction to data visualizations, statistical concepts, and the programming language, R. In the last half students completed a final project by analyzing a dataset using R. Students presented their projects in an expo-style format in which parents and family were invited to attend. The learning goals of this course were to (1) use programming and statistics as tools to analyze, visualize, and make claims about data and (2) reflect on the social and ethical implications of algorithm bias. In the course, students engaged in four algorithm bias activities. The first activity was an introductory discussion about Amazon's shopping algorithms. The second activity was an embodied activity that simulated Amazon's biased hiring algorithm that discriminated against women applicants. In the third activity, students watched a video about algorithm bias, which included discrimination issues with facial and photo recognition technology. In the final activity, students used Google image search and discussed the results. This initial analysis focused on the final algorithm bias activity and discussion. We segmented the conversation into 72 lines of turns of talk, developed a coding scheme based on Philip and colleagues framework on big data for democratic participation, and coded each line. Then, we divided the discussion into four topical sections, or stanzas, and used Epistemic Network Analysis (Shaffer & Ruis, 2017) to visualize the data.

Results and discussion

This analysis highlights the discourse of Ted, Alexander, and Pat, three participants with different contributions to the discussion. At the start of the discussion, Ted immediately introduced race and gender issues by stating "So

first of all besides Steve Harvey, 'game show hosts' are a bunch of *(hesitates)*, I mean, old White dudes. And if you look at 'nurse,' there's a bunch of women and a majority of them are White and if you look up 'doctor,' there's a few girls, but it's mostly men." He continued to describe the inequities in gender when searching for athletes. Other students joined the discussion, noticing that there were more men than women in the image search for basketball and tennis. Then, Ted posed a question: "Well, here is the question: Would you rather have accurate results when you Google physics professor and it's all old, White men or would you rather have it be very diverse but that's not the majority of physics professors. So that's the question: *would you rather have it be accurate or would you rather have it be diverse?*" Ted's question powerfully framed the remaining discussion as a dichotomy between "accuracy" and "diversity." Alexander shared a concern about which search results could be diversified: "But how's it supposed to know what topics don't need balancing out? Because like if you say give me a red apple, that would totally ruin the results by saying something like that… it wouldn't make the results mean anything." Alexander claimed that the algorithm would not know which topics to appropriately diversify. He argued diversifying image results for an apple would be meaningless. The conversation shifted to bias that impacts people. For example, Pat, an Asian-American male, shared this story: "So a while ago, the iPhone X, the face recognition, so this kid, he was Chinese, his mom said the face recognition was for her face but when he put it up to his face, it recognized him and said they were the same thing." In response, another student, expressed "That's messed up," indicating that the situation was unsettling or, potentially, unjust.

The ENA results display summary visualizations of each key student's contributions to the conversation in terms of a discourse network (Figure 1). The networks revealed that students made connections across the three categories of content/discursive fluency, motivated use of content, and politics of knowledge, but did so in different ways. Pat focused on racial bias and social justice, Alexander focused on the structure of the algorithms and machine learning concepts, and Ted had an overall balanced network.

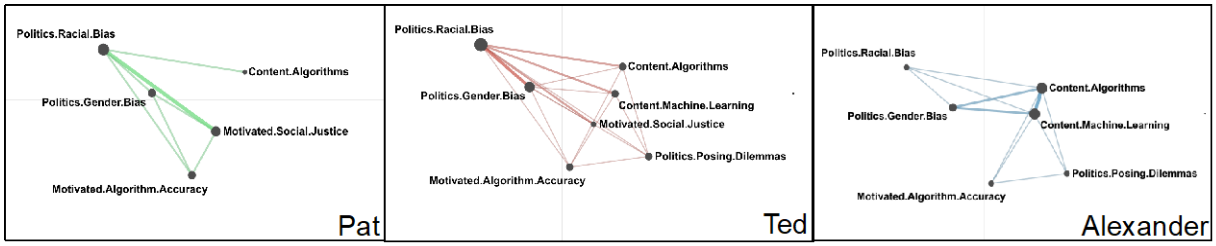

Figure 1. Discourse networks for three key students: Pat (green, focus on racial bias and social justice), Alexander (blue, focus on machine learning and algorithms), and Ted (red, overall distributed network.)

The implications of this initial work are that through structured activities, youth can engage in complex, ethical big data dilemmas that are part of a broader social conversation. This study is a foundation for broadening investigations on how diverse populations of learners make sense of algorithm bias and for developing more robust learning frameworks to inform future teaching and learning in areas of critical data literacies and learning.

References

Jiang, S., & Kahn, J. B. (2019). Data Wrangling Practices and Process in Modeling Family Migration Narratives with Big Data Visualization Technologies. *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings*. 13th International Conference on Computer Supported Collaborative Learning, Lyon, France.

Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism* (1 edition). NYU Press.

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1 edition). Crown.

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A Framework for Learning About Big Data with Mobile Technologies for Democratic Participation: Possibilities, Limitations, and Unanticipated Obstacles. *Technology, Knowledge and Learning*, *18*(3), 103–120. https://doi.org/10.1007/s10758-013-9202-4

Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making Space for Place: Mapping Tools and Practices to Teach for Spatial Justice. *Journal of the Learning Sciences*, *26*(4), 643–687. https://doi.org/10.1080/10508406.2017.1336440

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. *Handbook of Learning Analytics*.