WAKING UP TO MARGINALIZATION: PUBLIC VALUE FAILURES IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

Thema Monroe-White (Berry College)
Brandeis Marshall (DataedX)
Hugo Contreras-Palacios (Berry College)
INTRODUCTION

- AI is only one phase in the data science ecosystem — from the AI innovators and architects (in CS, Math and Statistics) to the AI technicians and specialists (in CS, IT and IS)

- *how we learn and what we learn is highly dependent on who we are as learners*

- *benefit hoarding: when “goods and services are not distributed equally”*

- *provider availability: the “scarcity of providers when an essential good or service is needed”*

- We examine demographic disparities by race/ethnicity and gender within the information systems educational infrastructure from an evaluative perspective

- We show bias creep including the situational exclusion of individuals from access to the broader knowledge economy
RESEARCH PROBLEM

HOW MIGHT WE BRING SUFFICIENT ATTENTION TO RACE/ETHNICITY AND GENDER (R/E&G) MARGINALIZATION IN DATA SCIENCE (DS) USING SOCIAL INCLUSION METRICS? IN WHAT WAYS MIGHT A QUANTITATIVE INTERSECTIONAL APPROACH HELP TO TELL THAT STORY? WHAT CAN BE DONE TO MITIGATE THESE DISPARITIES?
WE MEASURE MARGINALIZATION IN TERMS OF THE PROPORTION OF DS GRADUATE STUDENTS OF A GIVEN R/E & G GROUP ($DS_{GR}$) TO THE PROPORTION OF ALL GRADUATE STUDENTS ACROSS THE INSTITUTION ($I_{GR}$). WE CALCULATE AN INSTITUTIONAL PARITY SCORE $PS_{gr}$ AS THE LOG TRANSFORM OF THESE TWO RATIOS.

$$PS_{gr} = \log_2 \frac{DS_{gr}}{I_{gr}}$$
National NCES datasets of Master’s and Doctoral graduate students in IS, CS, Math and Statistics are used to create an “Institutional Parity Score” that calculates field-specific representation by race/ethnicity and gender in data science related fields.
In our study, we evaluate institutions with data science programs from 10 HBCU, 25 HSI and 254 PWI, representing a total of 68,949 graduate students and 6,839 faculty.
KEY FINDINGS

USING TWO PRIMARY SOCIAL INCLUSION METRICS (RATIOS AND INSTITUTIONAL PARITY SCORES), WE OBSERVE MAJOR DIFFERENCES IN SOCIAL INCLUSION BY R/E & G.
Women CS/IS, Math & Stats Graduate Students by race/ethnicity

WF: 4.28% more likely to see themselves in instructor roles

BF: 1.35% less likely to see themselves in instructor roles
HBCUs clearly outperform other types of institutions w/r/t Black women in DS related fields
INSTITUTIONAL PARITY & REPRESENTATION SCORES

Men CS/IS, Math & Stats Graduate Students by race/ethnicity

WM 25.96% more likely to see themselves in instructor roles

BM 1.95% less likely to see themselves in instructor roles
Both HBCUs and HSIs enroll more Black and Latinx graduate students respectively in CS, IS, Math & Stats graduate programs than PWIs.

HBCUs and HSIs together:

- Responsible for the lion’s share of Black and Latinx DS graduate students (both men and women).
- Responsible for a sizable share of Asian women DS graduate students in the case of both HBCUs and HSIs.
- Responsible for a sizable share Asian men DS graduate students in the case of HSIs.
**FINDINGS & INSIGHTS**

- White male graduate students are slightly “underrepresented” in DS fields according to our institutional parity score whereas Black men in DS on average have achieved near “parity”

- Question: What does it mean to be “underrepresented”?

- Answer: It depends on who your comparison group is.
  
  - If we were to compare the proportion of Latinx women in DS to their representation across the country or within a given state, then the representation score for Latinx women would be *much, much* worse.

- An institutional parity score is a valuable metric; however, it is insufficient for describing race/ethnicity and gender inequalities in academic institutions with data science programs.
LIMITATIONS

- Limited publicly available disaggregated faculty data by institution
- Non-US citizen data are not disaggregated by race/ethnicity
- Cross sectional view (2018) as opposed to trends over time
- Descriptive analyses
- Business school graduate student and faculty data not represented (i.e., I.S., MIS programs etc.)
- Other minoritized groups not represented: sexual orientation, immigrant status, etc.
KEY CONTRIBUTION

USING BOTH A NOVEL THEORETICAL FRAMEWORK AND METHODS FROM A CRITICAL QUANTITATIVE PERSPECTIVE, WE DESCRIBE AND ILLUSTRATE SOCIAL INCLUSION GAPS IN DATA SCIENCE AND PRESENT POLICY RECOMMENDATIONS THAT BRING US CLOSER TO SOCIAL INCLUSION GOALS.
CONCLUSION

- We need a combined quantitative and qualitative measures of
  - *participation* —assessing attraction strategies to DS
  - *access* —evaluating data skills attainment programs
  - *inclusion* —supporting retention activities
  - *representation* —developing equitable succession planning
- We provide a simple analytical approach for action-oriented scholars seeking to advance DEI in data science
- Highlighting the importance of meeting both the technical requirements and humanistic values essential to creating a more diverse data science workforce
THANKS FOR YOUR TIME & ATTENTION

tmonroewhite@berry.edu
brandeis@dataedx.com