Journal of Machine Learning Research 142:7–21, 2021 2nd Workshop on Diversity in Artificial Intelligence (AIDBEI)

Waking up to Marginalization: Public Value Failures in Artificial Intelligence and Data Science

Thema Monroe-White TMONROEWHITE@BERRY.EDU
Berry College

Brandeis Marshall BRANDEIS@DATAEDX.COM
DataedX

Hugo Contreras-Palacios HUGO.CONTRERASPALACIOS@VIKINGS.BERRY.EDU
Berry College

Editors: Deepti Lamba and William H. Hsu

Abstract

Data science education is increasingly becoming an integral part of many educational structures, both informal and formal. Much of the attention has been on the application of AI principles and techniques, especially machine learning, natural language processing and predictive analytics. While AI is only one phase in the data science ecosystem, we must embrace a fuller range of job roles that help manage AI algorithms and systems — from the AI innovators and architects (in CS, Math and Statistics) to the AI technicians and specialists (in CS, IT and IS). Also, it's important that we better understand the current state of the low participation and representation of minoritized groups that further stifles the accessibility and inclusion efforts. However, how we learn and what we learn is highly dependent on who we are as learners. In this paper, we examine demographic disparities by race/ethnicity and gender within the information systems educational infrastructure from an evaluative perspective. More specifically, we adopt intersectional methods and apply the theory of public value failure to identify learning gaps in the fast-growing field of data science. National datasets of Master's and Doctoral graduate students in IS, CS, Math and Statistics are used to create an "institutional parity score" which calculates field-specific representation by race/ethnicity and gender in data science related fields. We conclude by showcasing bias creep including the situational exclusion of individuals from access to the broader information economy, be it access to technologies and data or access to participate in the data workforce or data enabled-economic activity. Policy recommendations are suggested to curb and reduce this marginalization within in-formation systems and related disciplines.

Keywords: Data Science, Artificial Intelligence, Institutional Parity Score, Intersectional, Diversity, Graduate Education, HBCU

1. Introduction

The popularity of data science Venn diagrams (of which there are many!) (see Taylor (2016)) attempt to clarify its relationship to other more established fields (i.e., information systems, computer science and applied statistics); while others have championed the importance and relevance of big data and data science within the information systems (IS) discipline (Abbasi et al., 2016). Meanwhile, hundreds of academic and non-academic training programs have emerged across the country seeking to capitalize on this rapidly growing and increasingly popular field (Jafar et al., 2016). Many of these programs focus on establishing and enhancing learners' AI knowledge with the intent to leverage AI in operationalizing scalability. Be it text, audio and/or media, ma-chine learning, natural language processing and predictive analytics algorithms are foundational educational units that are being taught without a culturally responsive lens. History has shown us; however, that this proliferation will not benefit everyone equally exacerbating disparities between gender and racial/ethnic groups, (, NSF; Tatum, 2017) causing long-term harm to the social inclusivity of the field and as we argue here, public value failures.

Recent reports have highlighted the underrepresentation of women and minoritized groups in the data science workforce (Report, 2019; Duranton et al., 2020) in ways that resemble long-standing concerns by prominent scholars (Payton 2003). Despite data science being one of the fastest growing fields with high demand, women account for roughly 17 percent of professional data scientists (Burtch, 2019). The research brief by Noren et al. (2019) found with a little surprise that women make up just 26% of data science researchers and racially minoritized populations continue to be severely marginalized in data science research. At the core, access gaps to and interest in data instruction persists in marginalized communities further entrenching disparities in data-informed findings. To-date however, few studies have explored these disparities at the scale and scope of a landscape analysis (Sangiuliano and Cortesi, 2019; Windeler et al., 2018; Loiacono et al., 2016), and a small but growing number of studies have adopted quantitative intersectional approaches to target and elevate policies capable of mitigating these exclusionary trends in IS research (see Trauth et al. (2016) for an exception). The fields of AI have yet to design and release studies that are quantitative in nature while looking at race/ethnicity and gender inequities, even though at least one such study from Partnership in AI may be in progress. Information systems researchers have made greater strides in this social inclusion work; therefore, we lean on their models and frameworks to inform ours. Information systems anchors itself as a bridge between the information technology professionals who support technological operations and the users that leverage these tools for their own productivity end-goals. These encounters help us understand both sides of this continuum by revealing 1) how applications work and their relationship to other systems, 2) their capabilities and restrictions, when the application lifespan wanes and 3) how we transition to revised or new tech. Long-standing demographic data collection of students and instructors provides us with granularity information not recorded by other disciplines.

This demographic study contributes to the social inclusion research agenda by adopting equity-oriented frameworks from science policy and education literature and combining them with novel quantitative analyses of national publicly available demographic datasets. We then reveal and thereby critique the unique disparities faced by marginalized groups and link these findings to policies capable of mitigating these outcomes.

2. Related Literature

In this study we define data science as the ecosystem dedicated to the systematic collection, management, analysis, visualization, explanation, and preservation of both structured and unstructured data (Marshall and Geier 2019). We strengthen social inclusion research (Gorbacheva et al., 2019; Trauth, 2017), draw on the public value failure (Bozeman, 2007) and intersectionality (Hill Collins and Bilge, 2020; Crenshaw, 1991) literature to reveal more nuanced demographic disparities in academia. Public value failure (PVF) theory emerged from the science policy literature to counter economic narratives (i.e., market failures) that evaluate success in terms of dollars and cents and that dominate the practice of policy decision making. In particular PVF theory focuses on distributing resources more equitably to benefit the public (Bozeman, 2002, 2007) and seeks to evaluate outcomes based on answers the questions of who *benefits*? and who *is harmed*? This work calls for greater theoretical plurality (Stafford and Petter, 2019) while simultaneously uplifting critical approaches which have long been valued within the field (Hassan and Mingers, 2018; Lyytinen and Klein, 1985). In this way, this study attempts to reveal and critique the emerging discipline of data science in ways that challenge long-standing systems of power and white male hegemony in education (Ladson-Billings, 1998), prioritizing the experiences of racialized and minoritized data science, and by extension AI, instructors and learners for societal benefit

Public value failures are conceptualized as the absence of the public values or the provision of "rights, benefits, and prerogatives to which citizens should (and should not) be entitled" (Bozeman, 2007, p. 17) and which are essential to the human condition. For our purposes, the criteria most applicable within the data science educational system are *benefit hoarding* (when "goods and services are not distributed equally") and *provider availability* (the "scarcity of providers when an essential good or service is needed"). We define the *public* as anyone impacted by the products or outputs of data science workflows; and *citizens* are all those involved in the production and consumption of data science outputs including managers, programmers, researchers, and those who contribute to the production, dissemination and validation of data science products (Monroe-White and Marshall, 2019). Data science produces public value failures when members of minoritized groups are excluded from data workflows. The result includes harms to public welfare caused by biased recidivism risk (Dressel and Farid, 2018) and facial recognition models (Buolamwini and Gebru, 2018), ideologically white supremacist and sexist search engine outputs (Noble, 2018) among others. Ultimately, these failures affect those excluded from the data work and society as a whole.

Despite these devastating impacts, most of the scholarship on DS has focused on new methodologies, software or algorithms (the what) as opposed to DS practitioners (the who). This study prioritizes patterns within this lesser explored area by examining instructor-learner disparities by race/ethnicity and gender between DS providers (the institutions and faculty primarily responsible for preparing future data scientists for these roles) and learners (the students with potential to be data scientists) and is guided Therefore, this study is guided by the following research questions:

- RQ1. What are the disparities in representation among U.S. Black and Latinx instructors and learners within data science academic disciplines? To what extent do these disparities differ from those of whites and Asians?
- RQ2. How do these disparities vary by the type of U.S. academic institution?
- RQ3. What policy recommendations can be offered to create parity within the field as a whole?

3. Methods

Faculty and graduate student enrollment data along with institutional characteristics and data on DS graduate programs, are used to quantitatively illustrate intersectional differences in the types of disparities faced by DS learners and instructors from various race/ethnicity and gender demographic groups. National secondary datasets along with datasets on DS instructional offerings (i.e., degrees, certificates etc.) are mapped onto U.S. colleges and universities and analyzed using descriptive analyses and novel data visualization techniques.

3.1. Data sources and Variable Operationalization

The data field has yet to produce annual reporting on the enrollment, production, and employment of its learner and instructor population; therefore, by way of proxy variables, the count and proportion of data science faculty by race/ethnicity and gender were obtained from the 2018 Computing Research Association's (CRA) Taulbee Survey (Zweben and Bizot, 2019). The Taulbee Survey presents aggregated count and frequency data on tenure-track and non-tenure track faculty by race/ethnicity and gender for institutions offering computer science (CS), computer engineering (CE), or information (IS) PhD degrees. In 2018, survey data were collected from 164 institutions. The Taulbee survey defines Information (IS) programs as those consisting of Information Science, Information Systems, Information Technology, Informatics, and related disciplines with a strong computing component. Due to its focus on aggregate trends, however, the Computing Research Association's (CRA) Taulbee Survey does not provide a mechanism to unpack variations by institution or institutional characteristic.

Institutional characteristics as well as count data of potential data science graduate students by race, ethnicity and gender were obtained from the U.S. Department of Education National Centers for Science and Engineering Statistics 2018 Survey of Graduate Students and Postdoctorates in Science and Engineering (, NCSES). Data science (DS) graduate students are operationalized as those enrolled in a full time Master's or Doctoral degree program in Computer and Information Systems (i.e., GSS codes: 410, 411 and 412) or Mathematics and Statistics (GSS codes: 402 and 403). Institutions were also classified into one of three status types: historically Black college or university (HBCU); hispanicserving (HSI) or a predominantly white (PWI) institution. Lastly, programmatic data on Master's and Doctoral degree programs in data science were obtained from a third-party website (see http://datascience.community/colleges) in which institutions post details about their academic data science programs which includes information about the degree type (Bachelor's, Master's, Doctoral etc.) country of origin and state (if in the U.S.).

3.2. Data preparation

The data were prepared for analysis using the *pandas* and *numpy* packages in Python. Institution Name was used to join the datasets together. A combination of a fuzzy logic matching algorithm with a human-in-the loop component was used to reconcile institution names. All data visualizations were made with the free 2020.2 academic version of Tableau software. Initially, 471 institutions were retained representing 27 HBCU, 53 HSI and 391 PWIs; 463,250 graduate students 82,744 of whom are enrolled in computer and information sciences, mathematics and statistics programs (17.86%). Next, we restricted the dataset to those colleges and universities with verified data science programs. Institutions with DS Master's or Doctoral programs listed on the datascience.community site were all included in this list, followed by any additional institutions from the 2018 GSS survey with verified DS Master's or Doctoral programs. After removing institutions without DS graduate programs, 289 institutions remained. Lastly, institutions with fewer than 10 DS graduate students were removed leaving a total of 258 institutions representing 10 HBCU, 25 HSI and 254 PWIs; a total of 368,623 graduate students 68,949 of whom are in a DS field (18.70%).

3.3. Institutional Parity Score

In order to measure disparities by race/ethnicity and gender we first compute the ratio of the count of DS graduate students belonging to each gender and race combination to the total number of DS graduate students. This ratio is denoted as DS_{gr}. Likewise, we compute the ratio of all graduate students for each gender and race combination for all fields across the institution and divide it by the total number of graduate students enrolled at the institution. This ratio is denoted as I_{gr}. These two ratios are then used to create an institutional parity score denoted as PS_{gr}:

$$PS_{gr} = \log_2 \frac{DS_{gr}}{I_{gr}} \tag{1}$$

This approach matches the one developed by Tokita et al. (2015) for their representation metric, which they use to map changes in race/ethnicity and gender participation rates in undergraduate science, technology, engineering, and mathematics (STEM) fields. In this study, we adapt this formula and apply it to the interdisciplinary field of data science. To provide an intuitive explanation of this institutional parity score, the ratio of DS_{gr} and I_{gr} provides a measure of how likely a graduate student of a given race/ethnicity and gender (e.g., Latinx Male) is to encounter another graduate student of the same race/ethnicity and gender (e.g., Latinx Male) in a DS field as they would if they were a graduate student in another non-DS field at the same institution. Therefore, if this ratio is equal to one, then the chances of this student finding another student like themselves in the DS program is the same as that of the entire graduate school. If the ratio is smaller than one then the student is less likely to find a graduate student in DS like themselves than the entire graduate school. Likewise, if the ratio is greater than one, then the student is more likely to find a graduate student in DS like themselves than the entire graduate school. The logarithm acts as both a normalization and symmetrization tool. A score of 0 represents parity, where the proportion of DS graduate students belonging to a demographic group matches the proportion of graduate students belonging to that same demographic group across all graduate programs at the institution. Ultimately, this measure quantifies the degree to which certain demographic groups are likely to be over- or under-represented in data science within an institution.

CS and IS Faculty*	**Full**	**Associate**	**Assistant**	**Non-Tenure Track^**	*Race/ethnicity Subtotals* n (%)	**Grand Totals**
Black Female	4	11	13	14	42 (0.61%)	All CS/IS Female Faculty: **1,408**
Latinx Female	12	8	3	10	33 (0.48%)	
Asian Female	80	98	87	62	327 (4.78%)	
White Female	220	145	134	271	770 (11.26%)	
*Female Subtotals** (n)*	*316*	*262*	*237*	*357*	- -	
Black Male	15	16	16	24	71 (1.04%)	All CS/IS Male Faculty: **5,290**
Latinx Male	35	25	25	42	127 (1.86%)	
Asian Male	563	271	371	95	1,300 (19.01%)	
White Male	1,219	522	403	758	2,902 (42.43%)	
Male Subtotals (n)	*1,832*	*834*	*815*	*919*	- -	
Grand Total	**2,575**	**1,306**	**1,421**	**1,537**		All CS/IS Faculty^^: **6,839**

Note. *Grand totals represent counts of faculty by type for all race/ethnicity and gender categorizations, and therefore subtotal values will not sum to grant total calculations. ** Subtotals reflect sum of values for Black, Latinx, Asian and white classifications only and does not present values for Nonresident Alien, Indigenous American, Native Hawaiian/Pacific Islander, Multiracial and Ethnicity unknown or not reported categorizations. ^ Tenure-track faculty data reflect responses from 163 departments across 174 institutions; whereas non-tenure track data reflect responses from 154 departments. ^^This grand total value includes 141 CS/IS faculty whose race/ethnicity and/or gender are unknown. This value is also used at the denominator for all Faculty ratio calculations in the data visualization views.

Table 1: CS and IS Faculty counts by Race/Ethnicity and Gender

4. Analysis and Results

Descriptive analyses present disaggregated, intersectional ratio and institutional parity scores for eight demographic groups: Black, Latinx, Asian and white male and female potential DS faculty (*instructors*) and graduate students (*learners*) for the year 2018.

4.1. Instructors

In order to understand the landscape of potential providers of data science instruction, we present a summarizing view of 2018 Taulbee Survey results on CS and IS faculty across its 164 responding institutions by race, ethnicity and gender. Notably, most CS and IS faculty are Male (77.35%), white (53.69%) and Asian (23.79%) (see Table 1).

4.2. Learners

In this study, we operationalize data science (DS) learners via proxy variables provided in the NCSES GSS 2018 dataset. More specifically, we restrict our classification of potential DS learners to full-time graduate-level students (i.e., Master's and Doctoral) in computer and information systems or mathematics and statistics. In order to understand the comparative visualizations that follow, Figure 1 contains annotations to help explain the ratio and parity values for Black female DS instructors and learners. Notably, several institutions are not included in these views because they had no students of that particular demographic group in their 2018 Master's and Doctoral student enrollment numbers. As expected, these numbers are highest for Black and Latinx females and lowest for white males and females.

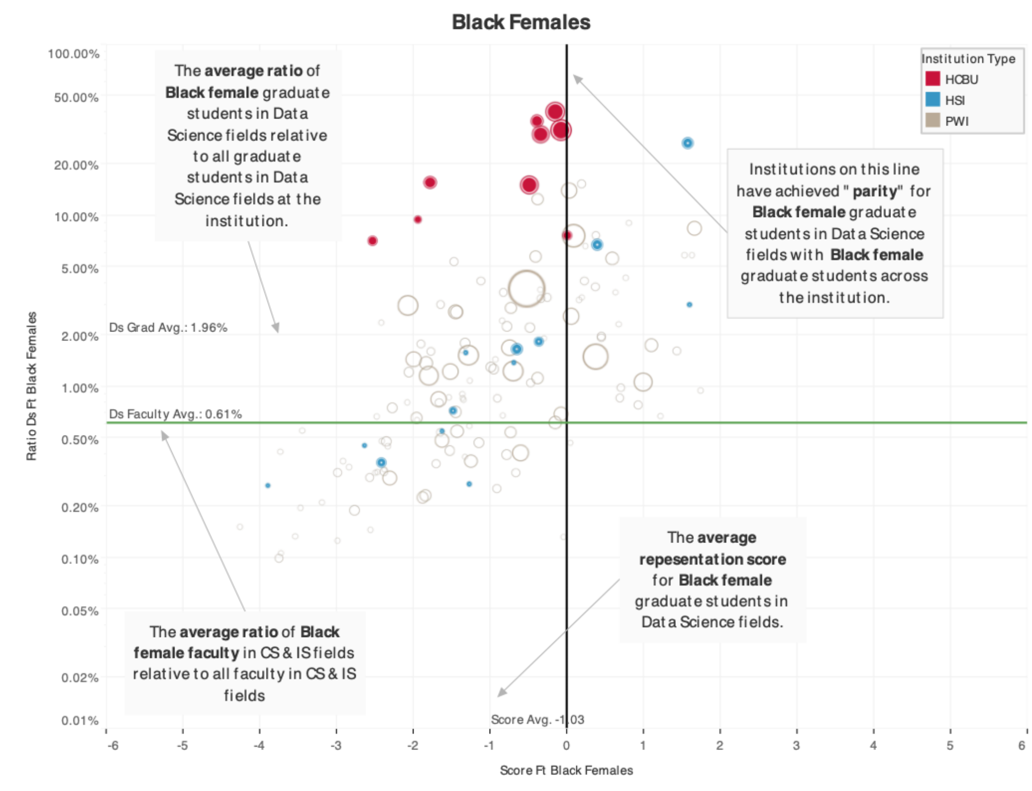

Figure 1: Black Female Master's and Doctoral Students (in CS, IS, Math or Statistics) at Institutions with Data Science Programs (labeled)

The x-axis on this scatterplot is the representation metric for Black female DS learners. The y-axis is the relative proportion of potential Black female DS learners compared to all potential DS learners at the institution. The size of the circle reflects the number of Black female Master's and Doctoral CS, IS, Math or Statistics students (i.e., potential DS learners), such that the larger the circle, the more potential Black female DS learners there are. The color of the circle indicates the type of institution: HBCU's are in red, HSIs in blue and PWIs in grey. The average proportion for Black female potential DS learners is 1.81% and is designated by the horizontal bar labeled "DS Grad Avg." The average proportion for Black female potential DS instructors is 0.61% and is designated by the horizontal green bar labeled "DS Faculty Avg." Lastly, the vertical bar labeled "Score Avg." designates the average representation score (-1.03) for Black female graduate students in DS fields. Together, these results indicate that on average Black female DS graduate students are underrepresented relative to other graduate-level disciplines at the institution. So much so in fact, that the chances of a Black female DS graduate student seeing another Black female DS graduate student at their institution is on average less than 2%, and the chances of seeing a Black female DS instructor is approximately one half of 1%. Figures 2 and 3 visualize this same set of relationships for potential DS learners across our four race/ethnicity (Black, Latinx, Asian, white) and two gender (Male, Female) demographic groups of interest in this study.

The visualizations in Figures 2 and 3 illustrate the severity of the representation problem among potential DS graduate students and faculty. Using an intersectionality perspective (Crenshaw, 1991), our findings suggest that over- and under-representation in instructors and learners of data science vary dramatically by both race/ethnicity and gender. Asian males and white males are overrepresented in both the instructor and learning category whereas Black and Latinx populations experience the widest negative gap between instructors and learners (i.e., there are proportionally fewer instructors than learners) and are significantly underrepresented as both instructors and learners. Table 2 summarizes in differences for our eight key demographic groups.

4.3. Instructor-Learner Ratio Comparisons

As Table 2 illustrates, the percentage point difference between DS instructors and learners varies substantially by demographic classification. Across the institutions in our sample, the share of Black female DS faculty is 1.35 percentage points *lower* than the share of Black female DS graduate students (1.96 - 0.61). This difference is worse for Black males where the share of Black male DS faculty is 1.95 percentage points lower than the share of Black male DS graduate students. The share of white female DS faculty on the other hand is 4.26 percentage points *higher* than the share of white female DS graduate students and this difference is substantially larger for Asian and white males, whose share of DS faculty is a whopping 15.31 and 25.96 percentage points higher respectively than their share of the DS graduate student population. These data tell us that Black and Latinx learners when they do enter CS, IS, Math or Statistics disciplines are less likely to see individuals like themselves in instructor roles which results in cascading implications for future career success (i.e., mentoring, employment, research assistantship possibilities etc.)

4.4. Institutional Differences

Both HBCUs and HSIs enroll more Black and Latinx graduate students respectively in CS and IS graduate programs than PWIs. Furthermore, without these institutions, the disparities in representation between DS learners and instructors would be even worse for these student populations. Moreover, when we remove PWIs from our aggregations and again examine average graduate student representation ratios, we find that HBCUs and HSIs together are not only responsible for the lion's share of Black and Latinx DS graduate students (both male and female); they also contribute to increases in relative share of Asian female DS graduate students in the case of both HBCUs and HSIs and Asian male DS graduate students in the case of HSIs (Table 2).

Figure 2: Female CS/IS, Math and Stats Graduate Students by Race/ethnicity (unlabeled)

Figure 3: Male CS/IS, Math and Stats Graduate Students by Race/ethnicity (unlabeled)

Race/ ethnicity & Gender	Institutional Parity Score	Avg. DS Faculty Ratio	Avg. DS Graduate Student Ratio**				Percentage Point Diff. in Instructor-Learner Ratios*
			All Inst.	*PWI only*	*HBCU only*	*HSI only*	*All Inst.*
Females							
Black	-1.03	0.61%	1.96%	1.18%	**21.48%**	1.98%	-1.35%
Latinx	-1.66	0.48%	1.11%	0.73%	0.81%	**5.04%**	-0.63%
Asian	-0.25	4.78%	2.08%	1.97%	2.96%	**4.78%**	+2.70%
White	-1.81	11.26%	7.00%	**7.51%**	0.27%	4.61%	+4.26%
Males							
Black	-0.03	1.04%	2.99%	1.93%	**31.03%**	2.43%	-1.95%
Latinx	-0.19	1.86%	2.70%	2.01%	1.19%	**10.04%**	-0.84%
Asian	+0.56	19.01%	3.70%	3.61%	2.01%	**5.27%**	+15.31%
White	-0.18	**42.43%**	**16.47%**	**17.57%**	3.45%	10.75%	**+25.96%**

Note. Highest percentages are highlighted in bold by column for the Institutional Parity Score; Average DS Faculty Ratio and All institution Average DS Graduate Student Ratio and percentage point difference values. Highest percentages are highlighted in bold by row for all other fields. **The National Centers for Science and Engineering Statistics (NCSES) does not provide publicly available disaggregated institutional faculty demographic data by race, ethnicity and gender, so calculating percentage point differences in instructor-learning ratios by institution type is not possible at this time.

Table 2: CS and IS Faculty counts by Race/Ethnicity and Gender

5. Discussion

Through the use of institutional parity scores, we observe public value failures in *provider* and *benefit* hoarding learners among Black and Latinx minoritized groups. We find that parity has yet to be achieved with varying degrees of over and underrepresentation. We describe these phenomena, detail their limitations and present potential policy recommendations to bring us closer to parity.

5.1. The "Parity" Paradox and P.A.I.R. Principles

White male graduate students are slightly "underrepresented" in DS fields according to our institutional parity score (PS_{gr}= -0.18); whereas Black Males in DS on average have achieved "parity" (PS_{gr}= 0.03). Meanwhile, white male representation exceeds that of Black males in average DS graduate student and DS faculty representation by 13.48 (16.47-2.99) and 41.39 (42.43 – 0.61) percentage points respectively. The question then becomes, what does it mean to be "underrepresented"? From a quantitative perspective, it depends on your comparison group is (i.e., your denominator). If we were to compare the proportion of Latinx Females in DS to their representation across the country or within a given state, then the Latinx female representation score would be much, much worse. Ultimately, while intersectional quantitative studies provide a useful starting point for problem identification and humanizing the experiences of minoritized students, we find that a single institutional parity score is insufficient for describing race/ethnicity and gender inequalities in academic data science institutions. Instead, combined quantitative and qualitative measures of *par-*

ticipation (defined as assessing attraction strategies to DS), *access* (defined as evaluating data skills attainment programs), *inclusion* (defined as supporting retention activities) and *representation* (defined as developing equitable succession planning) contribute to a more holistic understanding of intersectional, racialized and minoritized instructor and learner experiences in data science.

5.2. Limitations

This study does not include graduate student or faculty data on all demographic groups because of 1) missing data for individuals with "Race/ethnicity unknown" or not reported, 2) the inability to disaggregate "Nonresident Alien" classification by race/ethnicity and 3) extremely low student and/or faculty values for "Indigenous American" or "American Indian/Alaska Native", "Native Hawaiian/Pacific Islander", "Multiracial" or "Two or more races." The Taulbee survey also presents a limited view of the demographic landscape of CS/IS academic institutions. It consists only of PhD-granting institutions and subsequently excludes teaching colleges, community colleges and a vast majority of minority serving institutions (i.e., HBCUs, and TWIs) and does not include data on sexual orientation or non-binary gender identity classes. Likewise, this study currently presents a snap-shot view of the data science instructor-learner academic landscape as opposed to a longitudinal view of changes over time.

5.3. Policy Recommendations

1. *Who counts and who is counted counts.* Minoritized students in data science fields are significantly less likely to see themselves in the front of the classroom than their white and Asian peers. A *graduate student collective focused on Black and Latinx learners can help provide support where little assistance exists at a student's home institution.*

2. *Increase access to publicly available race and gender disaggregated faculty datasets.* The lack of publicly available data by institution on faculty race/ethnicity and gender reinforces forms of disinformation which seek to mask the extent of the problem for minoritized groups.

3. *Invest in anti-racist pedagogies.* Data instruction approaches needs revision to center culturally competent pedagogy, including social context and algorithmic accountability.

4. *Commit to greater institutional support.* HBCU's are closer to reaching representation parity. They should be serving as the primary institution on multi-institutional grants with parity tech infrastructure support.

6. Conclusion

This paper provides a ground truth for action-oriented scholars seeking to advance knowledge of diversity, equity and inclusion in data science and its adjacent fields of information systems, computer science, math, statistics. All of which leverage AI concepts, algorithms and tools. We recommend practical mitigation strategies for addressing the severe under-representation of minoritized groups in data science. This study adds plurality into AI research landscape by "introduc[ing] new theoretical vistas and fresh ideas to inform our worldview..." (Stafford and Petter, 2019) while simultaneously challenging "assumptions of heterogeneity (Kvasny et al., 2009)" and systems of power in order to understand the marginalization and exclusion of minoritized groups in these fields (Trauth, 2017; Payton and Berki, 2019). Ultimately, this work offers a framework to equitably meet the technical requirements and humanistic values that are essential to creating a more diverse data science workforce and provides clear actionable recommendations that address educational pathways to broaden participation in information systems and data science in particular.

References

Ahmed Abbasi, Suprateek Sarker, and Roger HL Chiang. Big data research in information systems: Toward an inclusive research agenda. *Journal of the association for information systems*, 17(2):3, 2016.

Barry Bozeman. Public-value failure: When efficient markets may not do. *Public administration review*, 62(2):145–161, 2002.

Barry Bozeman. *Public values and public interest: Counterbalancing economic individualism*. Georgetown University Press, 2007.

Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR, 2018.

Linda Burtch. The burtch works study: Salaries of data scientists & predictive analytics professionals. *Burtch Works, LLC. https://www.burtchworks.com/wp-content/uploads/2019/06/Burtch-Works-Study_DS-PAP-2019.pdf*, 2019.

Kimberle Crenshaw. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stan. L. Rev.*, 43(6):1241–1300, 1991.

Julia Dressel and Hany Farid. The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1):eaao5580, 2018.

Sylvain Duranton, Jörg Erlebach, Camille Brégé, Jane Danziger, Andrea Gallego, and Marc Pauly. What's keeping women out of data science. *Boston Consulting Group (https://www.bcg.com/publications/2020/what-keeps-women-out-data-science.aspx)*, 2020.

Elena Gorbacheva, Jenine Beekhuyzen, Jan vom Brocke, and Jörg Becker. Directions for research on gender imbalance in the IT profession. *European Journal of Information Systems*, 28(1):43–67, 2019.

Nik R Hassan and John Mingers. Reinterpreting the Kuhnian paradigm in information systems. *Journal of the Association for Information Systems*, 19(7):6, 2018.

Patricia Hill Collins and Sirma Bilge. Intersectionality. second edition. *Polity Press*, 2020.

Musa J Jafar, Jeffry Babb, and Amjda Abdullat. Emergence of data analytics in the information systems curriculum. In *Proceedings of the EDSIG Conference ISSN*, volume 2473, page 3857, 2016.

Lynette Kvasny, Eileen M Trauth, and Allison J Morgan. Power relations in it education and work: the intersectionality of gender, race, and class. *Journal of Information, Communication and Ethics in Society*, 7((2-3)):96–118, 2009.

G Ladson-Billings. Just what is critical race theory, and what's it doing in a nice field like education? *International journal of qualitative studies in education*, 11(1):7–24, 1998.

Eleanor Loiacono, Lakshmi S Iyer, Deborah J Armstrong, Jenine Beekhuyzen, and Annemieke Craig. AIS women's network: Advancing women in is academia. *Communications of the Association for Information Systems*, 38(1):38, 2016.

Kalle J Lyytinen and Heinz K Klein. The critical theory of Jurgen Habermas as a basis for a theory of information systems. *In Research Methods in Information Sysfems. E. Mumford, R. Hirschheim, G. Fitzgerald, Eds. North-Holland, Amsterdam*, pages 219–236, 1985.

Thema Monroe-White and Brandeis Marshall. Data science intelligence: Mitigating public value failures using pair principles. *Proceedings of the 2019 Pre-ICIS SIGDSA Symposium. 4. https://aisel.aisnet.org/sigdsa2019/4*, 2019.

National Center for Science (NCSES) and Engineering Statistics. Survey of graduate students and postdoctorates in science and engineering Fall 2018, https://ncsesdata.nsf.gov/gradpostdoc/2018, 2018.

Safiya Umoja Noble. *Algorithms of oppression: How search engines reinforce racism.* NYU Press, 2018.

Laura Noren, Gina Helfrich, and Steph Yeo. Who's building your AI? research brief, 2019. Link: https://www.obsidiansecurity.com/whos-building-your-ai-research-brief/.

National Science Foundation (NSF). Women, minorities, and persons with disabilities in science and engineering. special report NSF 17-310. national center for science and engineering statistics, arlington, VA, https://www.nsf.gov/statistics/2017/nsf17310/, 2017.

Fay Cobb Payton and Eleni Berki. Countering the negative image of women in computing. *Communications of the ACM*, 62(5):56–63, 2019.

Harnham Report. USA diversity in data and analytics: A review of diversity within the data and analytics industry in 2019. *(https://www.harnham.com/us/2019-usa-diversity-in-data-analyticsreport; accessed: September 12, 2019)*, 2019.

Maria Sangiuliano and Agostino Cortesi. Institutional change for gender equality in research: Lessons learned from the field. *Venezia: Edizioni Ca' Foscari - Digital Publishing. DOI http://doi.org/10.30687/978-88-6969-334-2*, 2019.

Tom Stafford and Stacie Petter. Our paradigm for paradigms in IS: How many times to the well? *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 50(3):8–11, 2019.

Beverly Daniel Tatum. *Why are all the Black kids sitting together in the cafeteria?: And other conversations about race*. Basic Books, 2017.

David Taylor. Battle of the data science venn diagrams. *KDNuggets News*, 2016.

Christopher K Tokita, William EJ Doane, and Brian L Zuckerman. Reframing participation in postsecondary STEM education with a representation metric. *Bulletin of Science, Technology & Society*, 35(5-6):125–133, 2015.

Eileen Trauth. A research agenda for social inclusion in information systems. *ACM SIGMIS Database: the Database for Advances in Information Systems*, 48(2):9–20, 2017.

Eileen M Trauth, Curtis C Cain, Kshiti D Joshi, Lynette Kvasny, and Kayla M Booth. The influence of gender-ethnic intersectionality on gender stereotypes about IT skills and knowledge. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 47(3):9–39, 2016.

Jaime Windeler, Stacie Petter, Kathy Chudoba, Emma Coleman, and Grace Fox. 2018 AIS community report: Diversity and inclusion in the AIS. Special interest group on social inclusion (SIGSI), 2018. Re-trieved from: https://aisnet.org/page/DiversityInclusion.

Stuart Zweben and Betsy Bizot. 2018 CRA taulbee survey. *Computing Research News*, 30 (5):1–47, 2019. https://cra.org/resources/taulbee-survey/.