Social Exclusion in Data Science: A Critical Exploration of Disparate Representation in Higher Education

Thema Monroe-White
Berry College, tmonroewhite@berry.edu

Dr. Brandeis Marshall
DataEdX, brandeis@dataedx.com

Dr. Hugo Contreras-Palacios
Berry College, hugo.contreraspalacios@vikings.berry.edu

Social Exclusion in Data Science: A Critical Exploration of Disparate Representation in Higher Education

Completed Research

Thema Monroe-White, PhD
Berry College
tmonroewhite@berry.edu

Brandeis Marshall, PhD
DataedX
brandeis@dataedx.com

Hugo Contreras-Palacios, PhD
Berry College
hugo.contreraspalacios@vikings.berry.edu

Abstract

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. As a society, we adjust to technological innovations at different rates based on our level of exposure to and investment in continued learning. However, *how* we learn and *what* we learn are highly dependent on *who* we are as learners. By examining the intersectional race/ethnicity and gender representation within the higher education infrastructure and applying the individual differences theory of social inclusion to national datasets of graduate students in computer and information sciences, mathematics and statistics we create a disparate representation metric for data science-related fields. Findings suggest intersectional variation in access to the broader information economy via systematic patterns of exclusion across institutions of higher education. Finally, recommendations are suggested to curb this marginalization.

Keywords

Intersectionality, representation, critical race theory, individual differences theory, data science.

Introduction

Recent reports have highlighted the underrepresentation of women and minoritized groups in the data science (DS) workforce (Harnham Report, 2019; Duranton, et al., 2019) in ways that resemble long-standing concerns by prominent information systems (IS) scholars (Payton & White, 2003). The DS workforce is comprised of many IS scholars, professionals, and practitioners as both fields are closely related yet distinct. IS traditionally sits at the intersection of technology, business, and strategy with a focus on system design, analysis, and strategy of information dissemination. DS is an "ecosystem dedicated to the systematic collection, management, analysis, visualization, explanation, and preservation of both structured and unstructured data" (Marshall & Geier, 2019) that devises approaches to handle data's navigation and manipulation within and across software and hardware systems. IS is powered by data — with the assumption that the data is accurate, complete, reliable, relevant, and timely. However, with the avalanche of data, we can no longer take for granted the reliability of data in our IS structures.

Subsequently, we have reached a critical juncture in the IS field to better understand data analytical techniques, as developed in the field, from a design science perspective. Scholars have attempted to clarify this relationship to other more established fields (i.e., information systems, computer science and applied mathematics and statistics); while others have championed the importance and relevance of DS within IS (Agarwal & Dhar, 2014; 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). However, we know that this proliferation will not benefit everyone equally, exacerbating disparities between gender and racial/ethnic groups, (NSF, 2017) causing long-term harm to the field. However, by coupling DS with IS theoretical framings we achieve a fuller application and management of data analytics in business settings from an organizational and humanistic perspective.

In this paper, we critique the emerging discipline of DS in ways that challenge long-standing systems of power and white male hegemony in education (Harris, 1993; Ladson-Billings, 1998; Taylor et al., 2020; Diggs, 2021), prioritizing the experiences of racialized and minoritized students for broad societal benefit. Subsequently, we leverage the social inclusion (SI) and higher education literature to reveal disparities within DS educational pathways. DS is one of the fastest growing fields with high demand, yet, women account for roughly 17 percent of professional data scientists (Burtch, 2019). The research brief by Noren, Helfrich and Yeo (2019) found that women make up just 26% of AI researchers, and racially minoritized populations continue to be severely marginalized.

At the core, access gaps to and interest in data instruction persists in marginalized communities further entrenching disparities in data-informed findings. A failure to diversity the data workforce will lead to the proliferation of increasingly biased algorithms with the vast majority of data harms affecting marginalized populations (Benjamin, 2019; Monroe-White, 2021 also see Broussard, 2018; Buolamwini & Gebru, 2018; and Noble 2018). To-date however, few studies have explored these disparities at the scale and scope of a landscape analysis (Loiacono, et al., 2016; Windeler, et al, 2018; Sangiuliano & Cortesi, 2019), and a small number of papers have adopted quantitative intersectional approaches to target and elevate policies capable of mitigating these exclusionary trends in IS (see Trauth et al., 2016 for an exception). This study contributes to the SI research agenda by adopting equity-centric IS frameworks and applying them to higher education institutions. We leverage novel quantitative analyses of national publicly available demographic datasets; identify the unique racial/ethnic and gender disparities faced by marginalized groups in data instruction and then link these findings to policies capable of mitigating these outcomes.

Theoretical Background

SI in IS Research

The term social inclusion has contested meaning (Gidley, et al., 2010). Scholars have leveraged the term to shed insight on macro-level issues of poverty alleviation (Sen, 2000), economic performance (Oxoby, 2009) and work integration (Nyssens, 2008) to micro-level topics of sense of belonging (Abrams et al., 2004), and individual identity (i.e., national, ethnic, gender etc.) (McCrone & Bechhofer, 2008). Although there appears to be a lack of consensus on a single definition of the term social inclusion, the concept of social *exclusion* has a relatively rich background in the academic literature. Furthermore, an exploration of social inclusion research within the IS literature reveals a robust yet emerging literature pertaining to sources of variation in IT career interest and persistence based on one's identity (Kvasny, et al., 2009; Gorbacheva et al., 2019). As noted by Trauth and Howcroft (2006), "[s]ocial exclusion arises from social inequalities and in every population, there are groups who remain underrepresented in the information technology profession and underserved by the ICTs it develops." Trauth's *Individual Differences Theory of Gender and IT*; provides a framework capable of explaining barriers and biases affecting full participation of women in the information technology (IT) workforce. More specifically, it allows for a critical, intersectional perspective that is often overlooked or underspecified in most IS research studies (Hassan & Mingers, 2018). Intersectional approaches enable researchers to emphasize women's racialized and gendered experiences by explicitly situating minoritized women as central actors in the identification of power struggles and social inequalities (Crenshaw, 1991; Brown & Gershon, 2017). According to Individual Differences Theory, this variation is best expressed in terms of three overarching constructs: *individual identity*, *environmental influences,* and *individual influences*. The *individual identity* construct can be summarized in terms of demographic characteristics; and the *individual influences* construct relates to individual personality traits, personal experiences and social networks that "push back against attitudes, biases, and other barriers to one's pursuit of an IT career" (Trauth, 2017; p. 12). Alternatively, *environmental influences* address broader societal and organizational institutions (i.e., norms, laws, and policies), attitudes and behaviors) that affect an individual's identity. Given the complexity of the latter construct, it is further composed of four subconstructs: *culture* (i.e., attitudes, behaviors including stereotypical views of women in IT fields), *economy* (i.e., size of a localized IT economy based on wages and cost of living), *policy* (i.e., federal, local and organizational laws and policies that affect a women's ability to work) and *infrastructure* (e.g., child-care and transportation). Ultimately, the *Individual Differences* theoretical framework enables researchers to explore a wide array of alternative solutions to gender disparities in the IS profession. However, as the field of IS becomes increasingly intertwined with big data, data analytics, and artificial intelligence scholarship, it is imperative that the implications and harms associated with these disparities be expressly outlined and delineated. By explicitly outlining patterns of social exclusion ascribed to the lack of representation of minoritized women in the field, scholar-activists are better equipped to shape and influence policies aimed at mitigating future harms to the field and broader population.

We follow Trauth's guidance and explore the *identity* and *environmental influences* dimensions of individual differences theory to explore topics pertaining to "[i]ntersectionality and underrepresented groups–as users or as IS professionals" (Trauth, 2017; pg. 13) and "[t]he operation of socio-economic privilege in a society on the exclusion of certain groups from equal participation in the information society" (Trauth, 2017; pg. 14) albeit with a narrowed focus on the infrastructural higher education landscape. Furthermore, given the relative dynamism of IS as a field, rapid growth in training programs and hiring of professionals within traditional IS programs, and DS's unprecedented impact on our social institutions, our study is centered on the identity and environmental influences that ultimately shape the experiences of minoritized women in graduate related DS programs.

SI in Higher Education

Social inclusion, with its varied meanings and interpretations, has facilitated a wide-variety of research streams within higher education literature including improving campus climate for minoritized groups (Hurtado & Ponjuan, 2005) to refining and enhancing anti-discrimination policies (Collins & Bilge, 2020). The extensive and growing scholarship on the racialized and gendered experiences of higher education STEM (science, technology, engineering/computer science and mathematics) students (Xu, 2008; Riegle-Crumb & King, 2010; Landivar, 2013; Su & Rounds, 2015; Saw et al., 2018; Carter et al., 2019) lends additional support for the importance of this line of inquiry within the IS and DS fields in particular. Furthermore, inclusive experiences can vary both between and within higher education institutions. Faculty, students, and staff can experience varying levels of inclusivity by college, school, department, program, and classroom environment. Not to mention, the individual experiences of stakeholders from minoritized groups can vary dramatically by social identity characteristics (i.e., race/ethnicity, gender- and sexual identity, class, disability status etc.) (see Archer et al., 2005; Basit & Thomlinson, 2012; Pidgeon, 2016; Morán et al., 2019); thereby necessitating frameworks which allow for more personalized notions of inclusion and exclusion. Likewise, all U.S. institutions of higher education (HEIs) are not created equal. HEIs can vary dramatically with respect to their mission and focus which can have drastic implications on the long-term composition of the faculty, staff, and student body. Historically Black Institutions (HBCUs) for example are defined by the Higher Education Act of 1965 as "institutions of higher learning established before 1964, whose principal mission was then, as is now the education of Black Americans." Their historical roots originated in the mid-19th century, after the U.S. Civil War, to meet the needs of Black students (USDoE, 1991), prior to this date U.S. laws prohibited the formal and informal education of Black people. Meanwhile, Historically or Predominantly White Institutions (PWIs) are described as "institutions of higher learning in which white students account for 50% or greater of the enrollment" (Brown II & Dancy II, 2010). The mid-1600s through the mid-1700s constituted the founding era of the first eminent PWIs (i.e., Harvard, William and Mary, Yale) where education was primarily relegated to the U.S. northeast, serving upper class members of society, and modeled closely after Western European institutions (i.e., Oxford and Cambridge). Hispanic-Serving Institutions (HSIs) on the other hand are a recent creation, formerly established in 1992 as a designation assigned to any institution that enrolls a large concentration of Latinx students. Specifically, HSI's are defined as "accredited, degree-granting, public or private, nonprofit colleges and universities with 25 percent or more total undergraduate Hispanic full-time enrollment (FTE)" (Nellum & Valle, 2015). There are currently over 470 two- and four-year institutions that meet this designation and hundreds more nearing this designation (Garcia & Taylor, 2017). Fundamentally, the social inclusion and social justice focus of HBCU's distinguishes it from the PWIs and HSIs in ways that facilitate the upliftment of Black students in ways that HSIs and PWIs do not. Also, by examining the demographic composition of graduate programs by race/ethnicity and gender across the three types of HEIs we are nuancing the conversation and extending it beyond one-size fits all models of social inclusion transformation within the higher education landscape for minoritized populations. This analysis in turn, helps to inform the intersectional IS social inclusion literature by deliberately questioning and theorizing disparities in demographic representation within the higher education landscape and by proposing policy solutions to mitigate them. Therefore, this study is guided by the following research questions:

- RQ1. To what extent are minoritized Black and Latinx graduate students represented within academic disciplines?
- RQ2. How does this representation vary by the type of U.S. academic institution (HBCU, HSI, and PWI)?

Quantifying SI

In their seminal article, Harrison and Klein (2007), define three types of within-unit diversity: separation, variety and disparity. In the present paper, our explicit emphasis is on the intersectional experiences of graduate scholars. Therefore, the type of diversity which most closely aligns with the conceptualization of social exclusion described above is that of *disparity*. Disparity (aka inequality or asymmetry) is defined as "differences in concentration of valued social assets or resources" (ibid, p. 1200). According to the authors', maximum disparity is reached when "one member of the unit outranks all others. He or she holds the lion's share if not all of a valued unit resource" (p. 1207).

In this paper we build on and extend this definition; and apply it to the concept of *representation*, which is particularly important within the field of higher education. Specifically, representation is narrowly defined as 'descriptive' representation (Mansbridge, 1999) where the genders, races/ethnicities in the room "…are in their own persons and lives in some sense typical of the larger class of persons whom they represent." The term representation also has ties to the fundamental information system theoretical literature in which "the essential purpose of an IS is to provide a faithful representation of some focal real-world phenomena, thereby assisting its users to track states and state changes (events) in the phenomena it represents" (Recker, et al., 2019, p. 735). Therefore, by combining the terms disparity and representation we highlight the relative importance of striving for parity (Monroe-White, et al., 2021), and from a uniquely IS perspective, the importance of continually evaluating and assessing the achievement of this goal (Reingold et al., 2020). In this sense, *disparate representation* captures the unevenness in the relative presence of particular race/ethnicity and gender groups that ultimately lead to the prioritization and benefit of some groups over others. Finally, Harrison and Klein (2007) caution against the use of aggregate measures of diversity, particularly when the global value is used to characterize individual differences. We agree, and present clear evidence that a failure to disaggregate by race/ethnicity and gender leads to inadequate assumptions of parity that can lead to the reinforcement of long-standing disparities within academic disciplines. Additionally, global measures of diversity are counter-productive to intersectional research and scholarship. Therefore, this paper contributes to the IS empirical literature by utilizing Trauth's individual differences theory and operationalizing the *disparate representation* metric to examine intersections of race/ethnicity and gender by higher education institution type (i.e., HBCU, HSI or PWI).

Methods

Graduate student enrollment data along with institutional characteristics of DS graduate programs, are used to quantitatively illustrate intersectional differences in the types of disparities faced by DS students by 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 a novel measure of disparate representation.

Data Sources and Variable Operationalization

The data for this study were prepared using the *pandas*, *numpy* and *seaborn* packages in Python. Python was used to pre-process the data, including data cleaning and data reconciliation as multiple data sources were joined together based on institution name. The *numpy* package (Harris et al., 2020) was used to transform data and the *pandas* package (McKinney, 2010) was used for data manipulation and analysis. Our original dataset contained 471 institutions (27 HBCU, 53 HSI, and 391 PWIs), 463,250 graduate students, 82,744 enrolled in computer and information sciences, mathematics, and statistics programs

(17.86%). After removing institutions without verified DS programs (i.e., DS programs were verified via the *datascience.community* site), 414 institutions remained: 17 HBCUs, 40 HSIs, and 357 PWIs.

Analysis and Results

To compare the representation of DS graduate students by gender and race/ethnicity within DS academic programs we perform three distinct sets of analyses. First, we present the representation landscape for race/ethnicity and gender group in our institution sample (see Table 1). This is important, as our analyses are only as valuable as our understanding of the current landscape of the DS graduate student population. Next, we present the design and use of the disparate representation diversity metric using the fractional contribution to the entropy as the primary function in which the proportion of a particular social group (i.e., race/ethnicity and gender) is compared to the overall diversity score. The strength of this approach is that the disparate representation metric indicates the relative share of the overall diversity contributed to by a particular race/ethnicity and gender group.

Intersectional Representation by Institution Type

An overarching view of the DS academic landscape reveals structural differences between HBCUs and HSI and PWI consistent with their missions and historical origins (Table 1). Black men and women make up the vast majority of the DS graduate academic programs at HBCUs, whereas white women and men comprise the majority (albeit unevenly distributed) of the HSI and PWI environment.

Institution Type	Number of Institutions	Women				Men			
		Asian	Black	Latinx	White	Asian	Black	Latinx	White
HBCU n (%)	17	9 (8.41)	89 (83.18)	6 (5.61)	3 (2.80)	7 (3.89)	154 (85.56)	4 (2.22)	15 (8.33)
HSI n (%)	40	170 (28.57)	40 (6.72)	138 (23.19)	247 (41.51)	304 (20.67)	62 (4.21)	398 (27.06)	707 (48.06)
PWI n (%)	357	1446 (24.15)	494 (8.25)	446 (7.45)	3602 (60.15)	2663 (18.51)	843 (5.86)	1169 (8.12)	9713 (67.51)

Note. The shaded cells represent the race/ethnicity and gender groups with the greatest representation of our eight demographic groups.

Table 1. Representation of CS and Information Science, Math, and Statistics graduate students by Gender, Race/Ethnicity, and Institution Type

Measuring Disparate Representation

This approach builds on the work developed by Tokita et al., (2015) and incorporates the metrics described by Harrison and Klein (2007). In particular, we address some of the shortcomings of the entropy metric described in Harrison and Klein (referred to as the Teachman, 1980 index) and the institutional parity score introduced in our earlier work by proposing a disparate representation metric that calculates the contribution of each gender/ethnicity combination to the overall entropy score of an institution or the fractional contribution to entropy. This fractionalized score ranges from 0 to 1, and it is proportional to $p*log(p)$ where p is the ratio of the count of students on a given gender/ethnicity class and the total number of students. Notably, this score also yields the same value for different distributions. For example, consider a two-value distribution with classes A and B, the distribution $[p_A, p_B]$ yields an entropy of $H = - p_A * \log(p_A) - p_B * \log(p_B)$ so the entropy is the same for distributions [0.4, 0.6] and [0.6, 0.4]. This formulation has the added benefit of being interpreted as a percentage, where the disparate representation metric explicitly considers the weight of each class on the total entropy. Specifically, while the computation of entropy involves a probability distribution as follows:

$$H = - \sum_{i=1,\ldots,N} p_i \log(p_i)$$

Our novel disparate representation metric is operationalized as the contribution of a particular race/ethnicity and gender demographic group (e.g., Black Women or Latinx Men) to the overall entropy score and is calculated as the ratio of the product of the probability distribution times the log of the distribution (a_i). This value tells us the relative share of the overall diversity score that is contributed to by a particular race/ethnicity gender group:

$$\alpha_i = -\frac{p_i \log(p_i)}{H}$$

Table 2 illustrates aggregated entropy and disparate representation scores for HBCUs, HSIs, and PWIs in our sample. Intuitively, the contribution to the entropy behaves similarly to the ratio of the race/ethnicity and gender group and when combined with the entropy metric of diversity, we find differences in DS graduate student retention by institution type.

Institution Type	Entropy		Disparate Representation Scores							
	STEM	DS	Asian		Black		Latinx		White	
			Men	Women	Men	Women	Men	Women	Men	Women
HBCU	2.91	2.82	3.6%	4.4%	19.6%	17.5%	2.3%	3.2%	6.3%	1.9%
HSI	3.01	2.72	8.3%	5.6%	2.6%	1.9%	9.8%	4.8%	13.6%	7.2%
PWI	2.73	2.37	7.3%	4.7%	3.1%	2.0%	4.0%	1.9%	16.2%	8.9%

Note. Disparate representation scores above the 75th percentile (>=.085) are highlighted in grey. Percentage values will not sum to 100 as percentages reflect values for Black, Latinx, Asian and white classifications only and do not present values for Nonresident Alien, Indigenous American, Native Hawaiian/Pacific Islander, Multiracial and Ethnicity unknown or not reported categorizations.

Table 2. Representation of CS and Information Science, Math and Statistics graduate students by Gender, Race/ethnicity, and Institution Type

HSIs have greater overall STEM diversity than PWIs and HBCUs, and HBCUs have greater DS diversity than HSIs and PWIs. However, in DS, HSIs are primarily upholding representation of white and Latinx males (13.6% and 9.8% respectively); while HBCUs maintain higher representation for Black men (19.6%) and women (17.5%). Furthermore, the drop in diversity between STEM and DS fields is least dramatic for HBCUs (a difference of 0.09) and most extreme for PWIs (a difference of 0.36). These data further suggest that PWI exclusionary practices in STEM are reinforced in DS, compounding disparities and reducing the relative representation of marginalized groups at these institutions.

Discussion

Our disparate representation metric demonstrates an unevenness in social inclusion across race/ethnicity and gender groups by institution type. In particular, HBCUs are clearly having sustained success (Winfield, et al., 2019). According to a UNCF Report, HBCUs which account for 3% of the higher education institutions, graduated 24%of all STEM-related bachelor's degrees earned by Black students in the United States (Saunders & Nagle, 2019). Our results reinforce this power and influence of HBCUs in diversifying the DS/IS workforce. HBCUs have higher overall DS entropy scores (i.e., greater diversity) and are particularly effective at ensuring higher levels of representation for Black and men women in ways that HSIs and PWIs are not. Similarly, while HSIs are doing a good job of propping up the number of Latinx men in DS, they have more work to do with respect to Latinx women. Finally, PWIs are doing poorly relative to their institutional counterparts with regard to social inclusion in STEM and DS. These findings have implications for higher education policy. In his 2018 doctoral dissertation, Boykin finds that Historically Black colleges and Universities despite being created as a tool for "restorative justice in the context of extreme social oppression" (p. 7) and having educated marginalized students of color for nearly 200 years, continue to be under-supported and undervalued. A failure to recognize the unique contribution and strengths of HBCUs to the broader social inclusion agenda would constitute a blatant disregard for these clear and present institutional leaders. PWIs and HSIs can learn from the example set by HBCUs from recruitment and retention of faculty, staff, and students, to pedagogical frameworks and course content aimed at achieving greater social inclusion in DS.

This study also has limitations worth mentioning. The 2018 NCSES dataset used in this study does not include graduate student 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) low student and/or faculty values for "Indigenous American" or "American Indian/Alaska Native", "Native Hawaiian/Pacific Islander", "Multiracial" or "Two or more races." Furthermore, NCSES disaggregated datasets currently do not include race/ethnicity data on non-binary gender identity classes, sexual orientation, or non-US citizens. Likewise, this study presents a snapshot view of the DS graduate student landscape; where future studies might consider a longitudinal view tracking changes in DS composition and race/ethnicity and gender representation over time.

Study Insights and Recommendations

- The disparate representation metric demonstrates that minoritized students in DS fields are significantly less likely to interact with individuals like themselves in the DS programs than their white and Asian peers. This can perpetuate a culture of isolation and marginalization that is worse than other STEM fields leading to even fewer DS faculty of color in the academy.
- Data instruction approaches need revision to center culturally competent anti-racist pedagogy, including social context and algorithmic accountability in curricular design in order to facilitate a welcoming and inclusive environment for students from marginalized populations.
- DS is inherently interdisciplinary, therefore, the lack of publicly available disaggregated datasets by race/ethnicity and gender for all students and faculty, including those in non-STEM disciplines (i.e., Business/Management) and non-US citizens (i.e., those currently classified as 'non-resident aliens' in NCSES datasets) reinforces disinformation and masking the extent of the problem for minoritized groups across DS relevant fields.
- HBCUs have a long-standing track record in graduating minoritized science and engineering majors. They should be serving as the primary institution on multi-institutional grants with parity technological infrastructure support, particularly for social inclusion efforts affecting marginalized students of color.

To further promote racial equity, fundamental concepts and strategies need to be operationalized within higher education institutions and professional development training programs. Anti-racist pedagogy and practices are also instrumental in amplifying inclusivity and sharing best practices in data curricula. These approaches must be grounded in a critical race and intersectional theory of education (Crenshaw, 1991; Ladson-Billings & Tate, 2006; Collins & Bilge, 2020), critical quantitative theory (Sablan, 2019), and culturally responsive pedagogy (Gay, 2002). Unfortunately, much of data instruction research and practices are carryovers from the STEM disciplines, which have a documented history that lacks racial literacy (Benjamin, 2019), and maintain a persistent source of exclusion for minoritized groups. Washington (2020), for example, argues for the need to require cultural competence as a focus in university computing departments nationwide. The broader responsibilities of global citizenship can be made part of the curricula design, development, and delivery fabric. Example materials that will enhance post-baccalaureate and professional development instruction in IS and DS include the 1619 Project Curriculum (https://pulitzercenter.org/lesson-plan-grouping/1619-project-curriculum), The Abuse and Misogynoir Playbook (D'Ignazio et al., 2021) and Native Land Digital (https://native-land.ca/).

Conclusion

This paper lays a foundation for action-oriented scholars seeking to advance social inclusion in DS by reviewing the current DS educational landscape and providing practical mitigation strategies for addressing the severe underrepresentation of minoritized groups. This study also adds plurality to IS research by "introduc[ing] new theoretical vistas and fresh ideas to inform our worldview…" (Stafford & Petter, 2019), alongside novel empirical approaches while simultaneously challenging "assumptions of heterogeneity" (Kvasny et al., 2009) and systems of power to understand the marginalization and exclusion of minoritized groups in higher education and IS related disciplines (Payton & Berki, 2019). Finally, it offers policy recommendations to equitably meet IS requirements alongside the humanistic values essential to creating a more diverse IT workforce, providing recommendations for racial equity in IS and DS in particular.

REFERENCES

Abbasi, A., Sarker, S., & Chiang, R. H. 2016. Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*2), 3.

Abrams, D., Hogg, M. A., & Marques, J. M. (Eds.). 2004. *Social psychology of inclusion and exclusion*. Psychology Press.

Agarwal, R., & Dhar, V. 2014. Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research, 25 (3),*443-448.

Archer, L., Hutchings, M., & Ross, A. 2005. *Higher education and social class: Issues of exclusion and inclusion*. Routledge.

Basit, T. N., & Tomlinson, S. 2012. *Social inclusion and higher education*. Policy Press.

Benjamin, R. 2019. Race after technology: Abolitionist tools for the new Jim Code. *Polity Press*.

Boykin, C. M. S. 2018. *Negatively Stereotyping Historically Black Colleges and Universities as an Intergroup Process*. Doctoral dissertation, UC Berkeley.

Brown II, C. M., & Dancy II, E. T. 2010. Predominantly white institutions. In K. Lomotey (Ed.), *Encyclopedia of African American education* (p. 524-526). SAGE Publications, Inc., https://www.doi.org/10.4135/9781412971966.n193

Brown, N. E., & Gershon, S. A. 2017. Examining intersectionality and symbolic representation. *Politics, Groups, and Identities*, *5*(3), 500-505.

Broussard, M. (2018). *Artificial unintelligence: How computers misunderstand the world*. MIT Press.

Buolamwini, J., & Gebru, T. 2018, January. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability, and transparency* (pp. 77-91).

Carter, D. F., Dueñas, J. E. R., & Mendoza, R. 2019. Critical examination of the role of STEM in propagating and maintaining race and gender disparities. *Higher education: Handbook of theory and research*, 39-97.

Crenshaw, K. 1991. Mapping the margins: Intersectionality, identity politics, and violence against women of color. Stanford Law Review, 43(6), 1241-1300.

Collins, P. H., & Bilge, S. 2020. *Intersectionality*. John Wiley & Sons. ISBN: 978-1-509-53969-7

D'Ignazio, C., Turner, K., & Wood, D. 2021. *The Abuse and Misogynoir Playbook*. MIT Media Lab. https://www.media.mit.edu/articles/danielle-wood-and-katlyn-turner-co-author-article-the-abuse-and-misogynoir-playbook-for/.

Diggs, S. N. (2021). The delusion of privilege. *Administrative Theory & Praxis*, 1-8.

Duranton, J., Erlenbach, J., Brégé, C., Danziger, J, Gallego, A, & Pauly, M. 2020. *What's Keeping Women out of Data Science?* Boston Consulting Group. (https://www.bcg.com/publications/2020/what-keeps-women-out-data-science.aspx)

Garcia, G. A., & Taylor, M. 2017. A closer look at Hispanic serving institutions. *Higher Education Today*.

Gay, G. 2002. Preparing for culturally responsive teaching. *Journal of teacher education*, *53*(2), 106-116.

Gidley, J. M., Hampson, G. P., Wheeler, L., & Bereded-Samuel, E. 2010. From access to success: An integrated approach to quality higher education informed by social inclusion theory and practice. *Higher Education Policy*, *23*(1), 123-147.

Gorbacheva, E., Beekhuyzen, J., vom Brocke, J., & Becker, J. 2019. Directions for research on gender imbalance in the IT profession. *European Journal of Information Systems*, *28*(1), 43-67.

Harris, C. I. (1993). Whiteness as Property. *Harvard Law Review*, *106*(8), 1707–1791. https://doi.org/10.2307/1341787

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Oliphant, T. E. 2020. Array programming with NumPy. *Nature*, *585*(7825), 357-362.

Harrison, D. A., & Klein, K. J. 2007. What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of management review*, *32*(4), 1199-1228.

Hassan, N. R., & Mingers, J. 2018. Reinterpreting the Kuhnian paradigm in information systems. *Journal of the Association for Information Systems*, *19*(7), 6.

Hurtado, S., & Ponjuan, L. 2005. Latino educational outcomes and the campus climate. *Journal of Hispanic Higher Education*, *4*(3), 235-251.

Jafar, M. J., Babb, J., & Abdullat, A. (2016). Emergence of data analytics in the information systems curriculum. In *Proceedings of the EDSIG Conference ISSN* (Vol. 2473, p. 3857).

Harnham Report. 2019. "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)

Kvasny, L., Trauth, E. M., & Morgan, A. J. 2009. 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.

Ladson-Billings, G. 1998. 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.

Ladson-Billings, G., & Tate, W. F. 2006. Toward a critical race theory of education. *Critical race theory in education: All God's children got a song*, *11*, 30

Landivar, L. C. 2013. Disparities in STEM employment by sex, race, and Hispanic origin. *Education Review*, *29*(6), 911-922.

Loiacono, E., Iyer, L. S., Armstrong, D. J., Beekhuyzen, J., & Craig, A. 2016. AIS Women's Network: Advancing women in IS academia. *Communications of the Association for Information Systems*, *38*(1), 38.

Mansbridge, J. 1999. Should blacks represent blacks and women represent women? A contingent" yes". *The Journal of politics*, *61*(3), 628-657.

McCrone, D., & Bechhofer, F. 2008. National identity and social inclusion. *Ethnic and racial studies*, *31*(7), 1245-1266.

McKinney, W. 2010, June. Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51-56).

Monroe-White, T. (2021, June). Emancipatory Data Science: A Liberatory Framework for Mitigating Data Harms and Fostering Social Transformation. In *Proceedings of the 2021 on Computers and People Research Conference* (pp. 23-30).

Monroe-White, T., Marshall, B., & Contreras-Palacios, H. (2021, September). Waking up to Marginalization: Public Value Failures in Artificial Intelligence and Data Science. In *Artificial Intelligence Diversity, Belonging, Equity, and Inclusion* (pp. 7-21). PMLR.

Morán, M. L., Gómez, L. E., Alcedo, M. Á., & Pedrosa, I. 2019. Gender differences in social inclusion of youth with autism and intellectual disability. *Journal of autism and developmental disorders*, *49*(7), 2980-2989.

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, 2017; https://www.nsf.gov/statistics/2017/nsf17310/

Nellum, C. J., & Valle, K. 2015. Government investment in public Hispanic-serving institutions.

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

Noren, L., Helfrich, G., Yeo, S. 2019. Who's Building Your AI? Research Brief. Link: https://www.obsidiansecurity.com/whos-building-your-ai-research-brief/

Nyssens, M. 2008. 6 The third sector and the social inclusion agenda. *The third sector in Europe: Prospects and challenges*, *8*, 87.

Oxoby, R. 2009. Understanding social inclusion, social cohesion, and social capital. *International Journal of Social Economics*.

Payton, F. C., & White, S. D. 2003, April. Views from the field on mentoring and roles of effective networks for minority IT doctoral students. In *Proceedings of the 2003 SIGMIS conference on Computer personnel research: Freedom in Philadelphia--leveraging differences and diversity in the IT workforce* (pp. 123-129).

Payton, F. C., & Berki, E. 2019. Countering the negative image of women in computing. *Communications of the ACM*, *62*(5), 56-63.

Pidgeon, M. 2016. More than a checklist: Meaningful Indigenous inclusion in higher education. *Social inclusion*, *4*(1), 77-91.

Recker, J., Indulska, M., Green, P., Burton-Jones, A., & Weber, R. 2019. Information systems as representations: A review of the theory and evidence. *Journal of the Association for Information Systems*, *20*(6)

Reingold, B., Widner, K., & Harmon, R. 2020. Legislating at the intersections: Race, gender, and representation. *Political Research Quarterly*, *73*(4), 819-833.

Riegle-Crumb, C., & King, B. 2010. Questioning a white male advantage in STEM: Examining disparities in college major by gender and race/ethnicity. *Educational Researcher*, *39*(9), 656-664.

Sablan, J. R. (2019). Can you really measure that? Combining critical race theory and quantitative methods. *American educational research journal*, *56*(1), 178-203.

Sangiuliano, M., & Cortesi, A. (Eds.). 2019. *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

Saunders, K. M., & Nagle, B. T. (2018). HBCUs Punching above Their Weight: A State-Level Analysis of Historically Black College and University Enrollment and Graduation. *Frederick D. Patterson Research Institute, UNCF.*

Saw, G., Chang, C. N., & Chan, H. Y. 2018. Cross-sectional and longitudinal disparities in STEM career aspirations at the intersection of gender, race/ethnicity, and socioeconomic status. *Educational Researcher*, *47*(8), 525-531.

Sen, A. 2000. Social exclusion: Concept, application, and scrutiny.

Stafford, T., & Petter, S. 2019. 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.

Su, R., & Rounds, J. 2015. All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in psychology*, *6*, 189.

Taylor, Barrett J., Cantwell, Brendan , Watts, Kimberly & Wood, Olivia. (2020). Partisanship, White Racial Resentment, and State Support for Higher Education. *The Journal of Higher Education* 91:6, pages 858-887.

Teachman, J. D. 1980. Analysis of population diversity: Measures of qualitative variation. *Sociological Methods & Research*, *8*, 341-362.

Tokita, C. K., Doane, W. E., & Zuckerman, B. L. 2015. Reframing Participation in Postsecondary STEM Education With a Representation Metric. *Bulletin of Science, Technology & Society*, *35*(5-6), 125-133.

Trauth, E. M., & Howcroft, D. 2006. Social inclusion and the information systems field: why now?. In *Social inclusion: Societal and organizational implications for information systems* (pp. 3-12). Springer, Boston, MA.

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

Trauth, E. M., Cain, C. C., Joshi, K. D., Kvasny, L., & Booth, K. M. 2016. 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.

US Department of Education (USDoE). 1991. Historically Black Colleges and Universities and Higher Education Desegregation. https://www2.ed.gov/about/offices/list/ocr/docs/hq9511.html

U.S. Department of Education (USDoE), National Centers for Science and Engineering Statistics (NCSES) 2018 Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS). 2018. https://www.nsf.gov/statistics/srvygradpostdoc/pub_data.cfm

Washington, A. N. 2020. When Twice as Good Isn't Enough. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. https://doi.org/10.1145/3328778.3366792

Windeler, J., Petter, S., Chudoba, K., Coleman, E., and Fox, G. 2018. 2018 AIS Community Report: Diversity and Inclusion in the AIS. Special Interest Group on Social Inclusion (SIGSI). Retrieved from: https://aisnet.org/page/DiversityInclusion

Winfield, L., Thomas, G., Watkins, L., and Wilson-Kennedy, Z.; Growing Diverse STEM Communities: Methodology, Impact, and Evidence ACS Symposium Series; American Chemical Society: Washington, DC, 2019. DOI: 10.1021/bk-2019-1328

Xu, Y. J. 2008. Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions. *Research in higher education*, *49*(7), 607-624.