The Role of Data and Artificial Intelligence in Driving Diversity, Equity, and Inclusion

Preeti S. Chauhan, IEEE
Nir Kshetri, University of North Carolina at Greensboro

Data and artificial intelligence hold promise for promoting diversity, equity, and inclusion. However, limited data availability, the biased nature of available data, and a lack of resources need to be overcome.

The year 2020 changed the world landscape in ways more than ever before. One of the differences that came about was the increased push to prioritize diversity, equity, and inclusion (DEI) at the workplace and in society at large. Data and artificial intelligence (AI) can be used to identify equality gaps and bring them to the forefront. According to U.S. census data,[1] in 2020, number of Caucasians in the working population grew steadily at every level of the corporate ladder, accounting for 85% of executives. This shows striking racial inequality at the executive level, where African Americans constitute only 2% of senior managers, while they make up 13.4% of the U.S. population. Similarly, only 3% of executives are Latin Americans, who constitute 18.5% of the population. These numbers show that the workforce inequality delta is deeply rooted and will require conscious efforts to correct. Another notable statistic is that African Americans have a median net worth that is 1/10th that of Caucasians.[2]

DEFINITIONS OF DEI

Diversity refers to all aspects of human difference, social identities, and social group characteristics, including but not limited to race, ethnicity, creed, color, sex, gender, gender identity, sexual identity, socioeconomic status, language, culture, national origin, religion/spirituality, age, and (dis)ability as well as military/veteran status, political perspective, and associational preferences. *Equity* describes fair and just practices and policies that ensure that all community members can thrive. Equity is different from equality in that equality implies treating

Digital Object Identifier 10.1109/MC.2022.3149017
Date of current version: 8 April 2022

0018-9162/22©2022IEEE

EDITORS
NORITA AHMAD American University of Sharjah; nahmad@aus.edu
PREETI CHAUHAN IEEE Reliability Society; preeti.chauhan@ieee.org


everyone as though their experiences were exactly the same. Being equitable means acknowledging and addressing structural inequalities—historic and current—that advantage some and disadvantage others. Equal treatment results in equity only if everyone begins with equal access to opportunities. *Inclusion* refers to a community where all members feel respected, have a sense of belonging, and are able to participate and achieve their potential. While diversity is essential, it is not sufficient. An institution can be diverse and uninclusive at the same time; thus, a sustained practice of creating inclusive environments is necessary for success.[3]

HOW CAN DATA AND RELATED TECHNOLOGIES DRIVE DEI?

Data and the latest analysis techniques can be used to drive deeper understanding of disparities at various levels of social and workplace settings. We can identify microtrends, generating real-time feedback instead of time-based surveys and making corrections as we go. A deeper dive into data can help locate hidden trends that otherwise might not be apparent. Failing to understand real trends can lead to false perceptions of and conclusions about DEI. Many companies believe they do not have a "significant" problem with diversity and inclusion.[4] However, without the use of deep analytics, some trends might not emerge at the first level of analysis. For instance, an examination of Facebook job listings found that 85% of the audience for supermarket cashier ads was women, while 75% of taxi company postings were aimed at African Americans.[5]

Appropriate data analysis techniques can help remove bias[6] that may inhibit diversity and inclusion. Human beings are inherently prone to biases (both conscious and unconscious) that may affect their judgment of diversity and inclusion. Machines, on the other hand, are dependent on data sets and algorithms developed by people but free of human biases. When AI and data analytics tools are well developed and deployed, they can identify and eliminate biases, especially unconscious ones that are much harder to uncover. AI can reveal hidden biases and alert the human resources and management representatives.

Data disaggregation[7] can help drive changes toward a more inclusive, diverse, and equitable society. When California's San Jose Office of Economic Development was trying to understand the business needs of the city's east side, it accommodated input from people of color. The data confirmed that large businesses got access to the Paycheck Protection Program to help them keep their workforce employed during the COVID-19 pandemic. However, small and microbusinesses with one to five employees were largely excluded. The problem was more pronounced in economically struggling neighborhoods, which were primarily Latino and Vietnamese. Further investigation revealed that there were undocumented businesses with language barriers and digital divides, which either did not trust the government to share information and didn't apply or did not know about the program. Once these trends emerged, the city began working with targeted community organizations, such as the Latino Business Foundation, to understand the companies' needs and offered solutions to create an incubator for the businesses.

DEI AT THE WORKPLACE

DEI is a continuous process and requires implementation at every level, including entry, middle management, and senior leadership. Timely and reliable data are at the core of driving a more diverse, equitable, and inclusive workforce. Companies that are successful in driving DEI regularly use data to establish and monitor metrics, track progress, and communicate to senior management, correcting course as needed. Data analytics tools, such as AI, can help improve decisions related to hiring, compensation, promotion, and retention, which can advance DEI. By analyzing historical information, AI can help reduce bias in the hiring process, recruit top talent, and identify conditions that may cause employees to quit. For instance, it can be used to recognize the optimum combination of skill sets for roles organizations are trying to fill. It can also discern bias-based language that may favor or offend certain groups based on age, ethnicity, and other factors. AI algorithms make decisions based on competency and skill sets that are required for jobs without knowing the applicants.[8]

AI is also being harnessed to create fairer opportunities for promotion and more equitable compensation. For instance, human resource managers and top executives often have unconscious ideas and feelings about an employee's readiness for promotion. AI algorithms can remove such bias and base decisions on test skills, aptitudes, and other factors that correlate with readiness for a position an employee is seeking. In this way, AI can help break barriers and glass ceilings for minority workers. It can also help organizations analyze market data to reduce and eliminate disparities in pay, promotions, and career advances that are associated with discrimination based on gender, race, and other conditions.[9]

AI can provide "predictive attrition" that enables businesses to identify factors that cause employees to quit. In 2019, IBM reported that it developed a "predictive attrition program" using

the Watson supercomputer and machine learning tools, which was able to forecast which workers would quit at a 95% confidence level. Likewise, people analytics and workplace planning software provider Visier's machine learning algorithms utilize hundreds of variables, such as job tenure, location, level, changes in pay, teams, the number of direct reports a person has, the length of time a person has been in a role, and the number of times a person's manager has changed, to predict outcomes, including job exits. The machine learning model, which was trained with real-world data sets involving employee behavior, can help companies gain insights so that appropriate interventions can be made to reduce attrition. For instance, companies could take measures to retain talented and skilled women and racial minorities.[10]

Data can serve as an impartial key to unlocking disparities faced by underrepresented populations in various sectors of society.

SUCCESS STORIES OF USING DATA TO DRIVE DEI

Data trends, when used effectively, can drive equity in career opportunities and advancement. One example is the South African social enterprise Harambee, which utilizes AI tools to help youth with job search opportunities. Harambee targets young adults from low-income backgrounds, with limited work experience and no job.[11] Since its founding, in 2011, its data-driven approach has helped young people find more than 240,000 opportunities. It uses Google's open source AI TensorFlow to interact with its network of more than 1 million young people and harnesses precise geographical attributes and preferential behavioral metrics to achieve its goals.[12] When the pandemic exposed women's employment vulnerability, including 400,000 lost jobs between 2020 and 2021, Harambee employed its data analytics to understand the core of the issue. The data revealed that almost 60% of young women were employed in wholesale and retail industries as clerks, support members, and sales staff. Those industries have not bounced back to prepandemic levels, exposing the market instability women face and the need for jobs which can withstand economic shocks.

On the other hand, men are more likely to work in sectors such as transportation and construction, which are faster to recover. Even in the case of formal economy jobs, due to their lower representation in managerial and highly skilled positions, women experienced twice the retrenchments that men did in 2020. Given these trends, there can be conscious efforts to place women in programs designed to grow jobs and reduce barriers to entry. One example is Harambee's partnership with CallForce and Vumatel in Soweto, which helps create township-based opportunities for women in global business services. When an organization intentionally strives for inclusive growth, it creates opportunities for people who were traditionally excluded. An example would be the plumbing sector, where Harambe partnered with the industry to lower barriers for young women. As a result, women now comprise more than 50% of the enrollees in the programs and are performing as well as or better than their male counterparts.

Another example where data has been used to drive meaningful DEI changes can be taken from the medical field. While clinical data sets, such as genomes, are biased against people of African ancestry, groundbreaking genome editing technologies focusing on this population have been developed thanks to increasing amounts of data. In December 2020, scientists successfully used gene editing to treat sickle cell disease, which is an inherited red blood cell disorder common among people of African ancestry. This approach is referred to as *precision medicine*, which relies on genetic information about a person or group for diagnosis and treatment. Sickle cell disease is associated with a lack of healthy red blood cells to carry oxygen throughout the body. To treat people who have it, scientists needed to identify and cut out a gene that causes the production of defective, sickle-shaped cells. Doing so requires sequencing the genome that makes up the deoxyribonucleic acid of people of African ancestry.[13]

CHALLENGES

Data can serve as an impartial key to unlocking disparities faced by underrepresented populations in various sectors of society. However, the interpretation of data is closely linked to the quality of the information and analysis techniques used. Some of the challenges of using data to drive DEI are discussed in the following.

The issue of multiple identification

A specific categorization can risk alienating those who identify with multiple groups. It is important that multiple and overlapping identifications are accommodated.[14] An example concerns the United Kingdom's Office for National Statistics (ONS). In 2021, in an effort to count Sikh members of the population, the ONS added a Sikh option to the ethnicity identification on the census. Some residents identified as Indian and Sikh. Because censuses do not allow ticking multiple options, the ONS reversed course. Similarly, there was a decision about whether to include Jewish as an option. During testing, it emerged that Jewish was included in either the Caucasian

or "other" sections. This presented Jewish as a white ethnicity and left out African American Jews, who already felt that their identities were unrecognized. These scenarios reveal that including more racial categories is not always the answer if multiple identification is not allowed.

Data divide and lack of resources for analysis

The data divide, including gaps in the availability of information for scientific research and decision making, adversely affects certain groups, such as minorities, women, and low-income individuals. In agriculture, for instance, index-based crop insurance programs have been promoted among small-scale farmers. In such models, preagreed insurance payouts are made based on the occurrence of clearly defined events.[15] For example, if the rainfall during a period is less than 50% of the average, a policyholder receives the maximum payout.[16] These schemes' effectiveness in reducing risk, however, depends on how well crop yields and farmer profits correlate with the index.[17] Data and measurement challenges limit small farmers' ability to benefit.

For example, to design an effective weather insurance policy, a significant number of weather stations is a necessary.[18] A criticism of East Africa's agricultural microinsurance provider for small farmers, Agriculture and Climate Risk Enterprise, is that its weather monitoring stations are primarily installed in heavily farmed areas in Kenya and Rwanda, leaving other regions with insufficient data. In addition, Canada's International Development Research Center found that the schemes' benefits have not materialized because the weather technology gathers data covering wide areas.[19] This provides general views about the effects of droughts and floods but fails to accurately measure rainfall at local levels. There have been cases in which satellite data indicated that an area had sufficient rainfall even though some farmers experienced crop loss due to microclimates. The farmers were not offered payouts, and some discontinued their policies.

Similarly, leveraging data to bridge gaps in health care needs to be carefully examined. Unfortunately, health records often leave out some important elements of patients' information pertaining to income, race, and ethnicity. This impacts downstream data analytics and decision making, often resulting in biased outcomes and the inequitable delivery of care. One example is undeveloped clinical data sets, such as genomes. Advances in human genomics make it possible to reconstruct population histories and identify genes that are likely to make people vulnerable to diseases. However, most genome studies have focused on people of European descent. In 2017, 19% of sequenced genomes belonged to people of non-European ancestry,[20] slightly increasing to 22% in March 2018.[21]

For instance, while India represents 20% of the world's population, only 0.2% of the fully mapped genomes relate to people of Indian origin. Likewise, people of African and Latin American descent and indigenous populations account for less than 4% of those the Genome Sequence DataBase.[22] Less than 3% of the genetic material used in global pharmaceutical research is from Africa. Africans and people of African descent are reported to be more genetically diverse than any other population.[23] The lack of genome data restricts underrepresented ethnic groups' ability to benefit from breakthroughs in human genomics that have been made in the past three decades. To address this, it is necessary to have holistic and standardized data collection to factor in the missing elements. Additionally, defining minimum sample sizes for statistically relevant data reporting will help tackle disparities.

Unavailability is not the only challenge to promoting DEI the use of data and AI. Many data and AI applications are computationally intensive. Training some machine learning algorithms requires expensive equipment, computing power, and costly methods. Small, developing countries are often unable to afford the price. Because of the lack of sufficient processing capacity, many government agencies are unable to use information they receive from their own satellites and from free sources.[24] Likewise, it was reported that some AI solutions can take as long as 12 h to run on typical computers in Kenya.[25]

Pitfalls of data aggregation

Data aggregation can lead to misinformation. In 2015,[26] Austin, Texas, was named the best place for families in the United States even though it was the most economically segregated city in the country. When a deep data dive was conducted to investigate, it was found that the Latino and African American populations were five to seven times more likely to live in poverty. Can we truly call Austin the best place for families, when there is such a high level of disparity and extreme poverty in certain communities? It is important to thoroughly explore data and understand microtrends to drive more effective decision making. Fairfax County, Virginia, offers another example of disaggregating data. Numbers for recidivism in the juvenile court system appeared to be low, but a closer examination revealed a different reality for youth of color. It was found that when white children entered the system, they were diverted—treatment never afforded to African-Americans and Latinos.

Data collection restrictions

DEI data are considered personally identifiable information, and there is a lack of consensus about collecting them. For example, the French and German governments do not collect data about race due to past horrors inflicted on minorities by the Nazis. However, without data, it is impossible to understand inequalities and drive mitigating actions. W.E.B. DuBois's team led the way, in 1890, to utilize

data visualizations to explain institutionalized racism to the world. With Black Lives Matter protests spreading all around the globe, the French and German governments' position on data collection came under pressure. A French government spokesperson, Sibeth Ndiaye, suggested that gathering racial data could help fight racism since it would enable the country to "measure and look at reality as it is." In Germany, an online survey, Afrozensus, was conducted in the third quarter of 2020 to gather information about discrimination faced by Germans of African descent. The survey received funding from the government but managed the data and storage on its own. Although these have not led to changes in the countries' stance on racial data collection, the efforts represent a good starting point to open discussions on the matter.

Data and AI can play important roles in addressing the lack of DEI in various settings. For instance, AI can help reduce human bias in decisions related to hiring, compensation, and promotion. By combining programs, such as predictive attrition, with human intervention, companies can take actions to retain marginalized groups. At the same time, limited data availability and the biased nature of available information have limited the potential of AI to promote DEI. For example, small farmers have not benefited from index-based insurance because of inadequate rainfall data. Likewise, while developments in clinical genomics can help diagnose and understand complex diseases and make decisions related to individualized medicine, most studies have focused on people of European descent. Hence, there is a need to refine practices and policies to address data collection barriers. A lack of resources, such as sufficient processing capacity, has also prevented efforts to use data and AI to benefit marginalized groups and promote DEI. Through time, these barriers are expected to decrease, which will help the private sector, governments, academic institutions, and other stakeholders make effective use of data and AI to advance DEI.

REFERENCES

1. A. Stahl, "What's to come in 2021 for diversity, equity and inclusion in the workplace," *Forbes*, Apr. 14, 2021. [Online]. Available: https://www.forbes.com/sites/ashleystahl/2021/04/14/whats-to-come-in-2021-for-diversity-equity-and-inclusion-in-the-workplace/?sh=34a1324c7f26
2. E. Kennedy, "Can data drive racial equity?" *Sloan Review*, Dec. 3, 2020. [Online]. Available: https://sloanreview.mit.edu/article/can-data-drive-racial-equity/
3. "DEI definitions," Univ. of Iowa. https://diversity.uiowa.edu/resources/dei-definitions (Accessed: Mar 6, 2022).
4. B. Banks, "Hiring more diverse people isn't enough': 4 things to really focus on to promote racial equity in 2021," CNBC, Dec. 16, 2020. https://www.cnbc.com/2020/12/16/hiring-more-diverse-people-isnt-enough-how-managers-ceos-can-promote-racial-equity-in-2021.html
5. M. Bogen, "All the ways hiring algorithms can introduce bias," *Harvard Business Rev.*, May 6, 2019. [Online]. Available: https://hbr.org/2019/05/all-the-ways-hiring-algorithms-can-introduce-bias
6. H. Zhang, S. Feinzig, L. Raisbeck, and I. McCombe, "The role of AI in mitigating bias to enhance diversity and inclusion," IBM. https://www.ibm.com/downloads/cas/2DZELQ4O (Accessed: Mar 6, 2022).
7. K. A. Dilday, "How data can create racial equity," Bloomberg, Mar. 24, 2021. https://www.bloomberg.com/news/articles/2021-03-24/how-data-can-create-racial-equity
8. E. Sweeney, "AI can help companies remove bias in the hiring process, speed up recruitment, and prevent employees from quitting, experts say," *Business Insider*, Nov. 9, 2021. [Online]. Available: https://www.businessinsider.com/how-ai-help-companies-recruit-keep-top-talent-2021-11
9. M. M. Robinson and S. Robinson, "How AI can help build a diverse workforce," People Management, Dec. 6, 2021. https://www.peoplemanagement.co.uk/voices/comment/how-ai-help-build-diverse-workforce#gref
10. N. Lewis, "Should I stay or should I go? Can AI predict that?" Society for Human Resource Management, Dec. 3, 2021. https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/ai-predicting-turnover-retention.aspx
11. O. Newton, "Preparing young people for work: Innovations from South Africa," FE News, Mar. 27, 2021. https://www.fenews.co.uk/featured-article/65849-preparing-young-people-for-work-innovations-from-south-africa
12. A. Paul, C. Jolley, and C. A. Anthony, "Reflecting the past, shaping the future: Making AI work for international development," USAID, 2019. www.usaid.gov/sites/default/files/documents/15396/AI-ML-in-Development.pdf
13. U. Edward-Ekpu, "Two Nigerian laboratories have taken big steps to boost genetics medicine in Africa," Quartz India, Dec. 14, 2020. https://qz.com/africa/1945960/nigerian-labs-acegid-54gene-boost-genetics-medicine-in-africa/
14. A. Mbembe, "The dividing line: How we represent race in data," Open Data Institute, 2020, https://theodi.org/article/the-dividing-line-how-we-represent-race-in-data/
15. A. Zhou, "How blockchain smart contracts are reinventing the insurance industry," Nasdaq, Oct. 29, 2021. https://www.nasdaq.com/articles/how-blockchain-smart-contracts-are-reinventing-the-insurance-industry-2021-01-29
16. "AgriBee(Cambodia)Plc," Agribee, https://agribee.co/home/category?parent_id=22&id=75&menu_id=0
17. J. W. Glauber, "Crop insurance reconsidered," *Amer. J. Agricultural Econom.*,

vol. 86, no. 5, pp. 1179–1195, 2004, doi: 10.1111/j.0002-9092.2004.00663.x.
18. S. L. Jorgensen, M. Termansen, and U. Pascual, "Natural insurance as condition for market insurance: Climate change adaptation in agriculture," *Ecological Econom.*, vol. 169, p. 106489, 2020, doi: 10.1016/j.ecolecon.2019.106489.
19. K. Njagi, "Kenyan farmers snap crops with phones to improve insurance payouts," Reuters, Oct. 11, 2019. https://www.reuters.com/article/us-climate-change-kenya-insurance/kenyan-farmers-snap-crops-with-phones-to-improve-insurance-payouts-idUSKBN1WQ0Q7
20. N. Parry, "Gene databanks ignore non-Europeans," Health Issues India, 2017. https://www.healthissuesindia.com/2017/08/30/gene-databanks-ignore-non-europeans/
21. G. Sirugo, S. M. Williams, and S. A. Tishkoff, "The missing diversity in human genetic studies," *Cell*, vol. 177, no. 1, pp. 26–31, 2019, doi: 10.1016/j.cell.2019.02.048.
22. G. Guglielmi, "Facing up to injustice in genome science," *Nature*, vol. 568, no. 7752, 2019, doi: 10.1038/d41586-019-01166-x.
23. "African genomics startup 54gene raises $25M to expand precision medicine capabilities," TechCrunch, Sep. 16, 2021. https://techcrunch.com/2021/09/16/african-genomics-startup-54gene-raises-25m-to-expand-precision-medicine-capabilities/
24. T. Anderson, "Launching your own satellite — the pros and cons," Sci Dev Net, 2009. https://www.scidev.net/global/features/launching-your-own-satellite-the-pros-and-cons/
25. D. Pilling, "AI in Africa healthcare falls short of potential," *Financial Times*, 2020. [Online]. Available: https://www.ft.com/content/90fa8f44-6847-11ea-a6ac-9122541af204
26. K.A. Dilday, "How data can create racial equity," Fuse Corps, Mar. 29, 2021. https://www.fusecorps.org/2021/03/29/how-data-can-create-racial-equity/

PREETI S. CHAUHAN is a technical program manager at Google, Sunnyvale, California, 94089, USA. She is a coeditor of the "Data" column for *Computer*. Contact her at preeti.chauhan@ieee.org.

NIR KSHETRI is the "Computing's Economics" column editor of *Computer* and a professor at the Bryan School of Business and Economics, University of North Carolina at Greensboro, Greensboro, North Carolina, 27412, USA. Contact him at nbkshetr@uncg.edu.

Digital Object Identifier 10.1109/MC.2022.3160106