Rebuilding an Online Assessment System Aimed at Identifying Individual Strengths and Weaknesses in Regards to Socioeconomic Power Differences

(Technical Paper)

Analyzing the Promise of Software Tools and Artificial Intelligence in Decreasing Bias and Establishing an Inclusive Work Environment

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, School of Engineering

Nuzaba Nuzhat

Spring, 2020

Technical Project Team Members

Amelia Hampford, Carl Zhang, Connor Yager, David Xue, Peter Felland, Sam Shankman, William Zheng

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Signature _____ Date _____

Nuzaba Nuzhat

Approved _____ Date _____

Ahmed Ibrahim, Assistant Professor, Department of Computer Science

Approved _____ Date _____

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction

In the past few decades, explicit and overt expressions of bigotry have seen an overall decline in the United States; however, the same cannot be said of their more implicit, aversive counterparts. (Bobo, Kluegel & Smith, 1997; Krysan & Moberg, 2016; see Figure 1). The difficulties are twofold: first, these behaviors are hard to uncover since people rarely wish to admit their presence (Pearson, Dovidio, & Gaertner, 2009), and second, even when the presence of such discriminatory beliefs and actions is clear, reducing it can become a mammoth task due to human beings' inherently biased nature (Lepore & Brown, 1997). To address this challenge, various for-profit and non-profit organizations have developed a multifaceted approach that consists not only of traditional training, workshops, assessments, and discussions, but also increasingly more and more software tools.

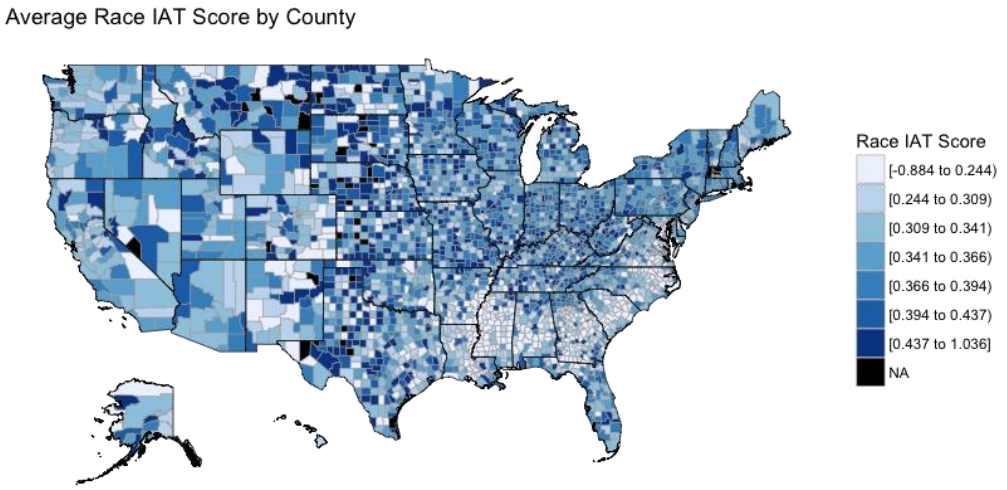

Figure 1: Average IAT score by US county. Lighter blue means less pro-white implicit bias. ("Project Implicit" n.d.)

For our technical project, we will be assisting one such organization, The Sum, augment their traditional diversity and inclusion offerings. Specifically, we will look into deploying a web application where users can sign up and take The Sum's Power of Difference Assessment (PDA) that will allow them to better understand how their behavior is conducive or deleterious to a

more diverse organization and society. Although there is already a version of the assessment deployed online, various issues related to usability, security, and scalability have rendered it unsuitable for further use. Our goal for this capstone project would be to build the system back from scratch to address these issues while adding features related to user administration and data analysis.

For my STS topic, I would like to investigate companies that pursue roughly the same goals as The Sum, but with a much greater dependence on technology. Categorized under the Diversity and Inclusion Technology (D&I Tech) industry, these companies promise revolutionary change by greatly reducing or replacing the work of biased humans with a mix of software systems and machine learning algorithms (Garr & Research, 2019). The STS research will analyze the positive and negative receptions these new tools have received, especially those that operate at the controversial intersection of artificial intelligence and bias-free hiring.

Technical Topic: Rebuilding an Online Assessment System Aimed at Identifying Individual Strengths and Weaknesses in Regards to Socioeconomic Power Differences

The Sum, led by Elliott Cisneros, is a Charlottesville, Virginia non-profit partnered with the Heather Heyer Foundation which promotes personal growth, skill development, and diversity. The goal of The Sum is to stand in solidarity with all people, no matter their background. The PDA gathers participants' demographics and asks a series of demographic-based questions. After taking the PDA, a report with results is generated and emailed to the participant. The results are categorized across demographics, areas of strength, and areas of growth (see figure 2). The results help reveal people's demographic biases. Those that take the PDA can meet with a consultant from The Sum to learn about their biases and how to

communicate better across demographics. There are paid, free, and organizational versions of the PDA with the only difference being the length of the consultation received ("The Sum", n.d.).

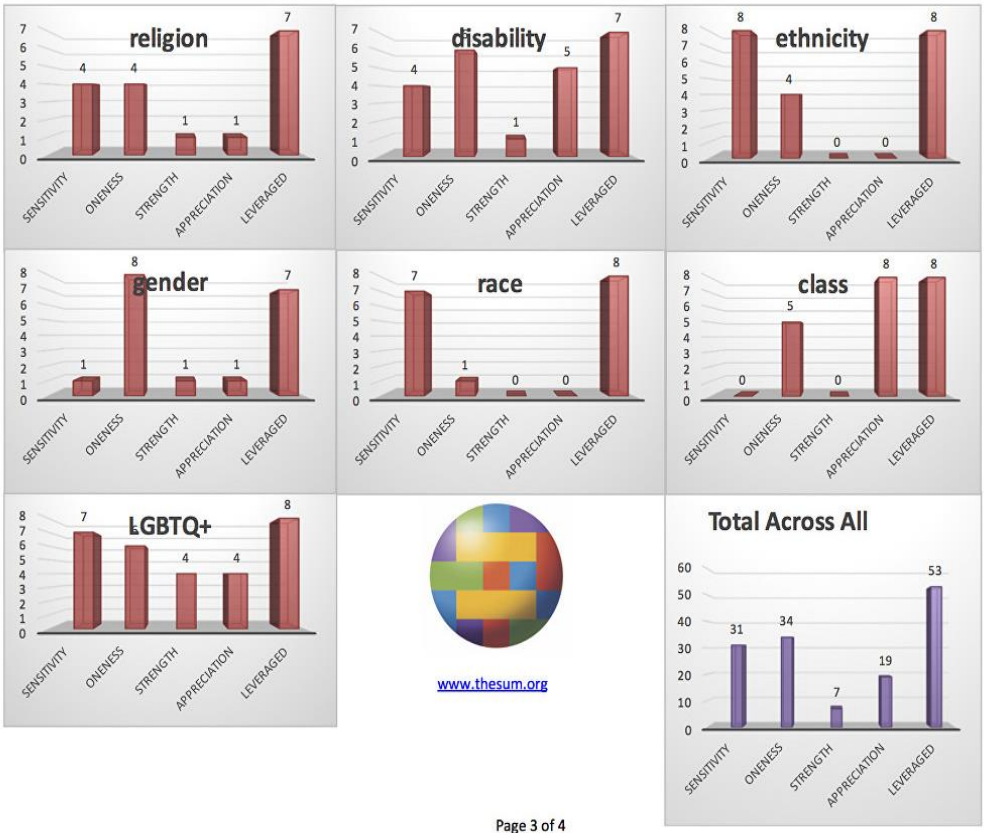

Figure 2: Sample result generated from the PDA. There are seven sociocultural locations: religion, disability, ethnicity, gender, race, class, and, LGBTQ+, and five power perspectives: sensitivity, oneness, strength, appreciation, and leveraged. To become an individual ready to create positive change in the world, you should have high scores for leveraged and nearly zero for everything else ("The Sum", n.d.).

The Sum already has an online PDA system in place. The current system allows users to take the PDA and schedule a consultation. However, the system is error-prone. The system improperly categorizes results and it requires someone at The Sum to manually generate reports and email them to users. As part of the report generation, categorizations are manually checked and corrected. This makes report generation time consuming and prone to human error. Although manually generating a report only takes a few minutes, the time from PDA completion to reports being emailed to users varies based upon availability at The Sum and can take up to 24 hours.

This methodology is not scalable and cannot support the upcoming UVA Department of Psychology study of 1,000 PDA takers. In addition to this, the current system does not detect a difference between assessment versions. It is also insecure and allows for URL manipulation.

The goal of this capstone project is to make a new PDA system. The new system should include all the features of the current system. The new system should correctly categorize results, generate reports, email reports to users, and detect which version of the PDA is being taken. For organizational and paid versions of the PDA, the new system should handle organizational access and payments correctly. The new system should also have security checks in place to prevent revisiting previously answered questions and URL manipulation.

To make the new system, requirements had to be gathered from The Sum. Requirements determine what features should be part of the new system and which features should be prioritized. Feature prioritization impacts the development timeline. Requirements help track development progress. Separating the work into requirements allows the team to determine who works on which features. Most importantly, requirements establish clarity between the capstone project team and The Sum for what is to be built.

Minimum requirements are to make a system where users can sign up with a valid email address, undergo email verification, select which version of the assessment to take, fill out user demographics, answer each question of the PDA, only view one question at a time, only answer questions in order, and have access to the separate consultant scheduling system. Users cannot change responses to previously answered questions. For the minimum requirements, assessment versions do not have to differ and The Sum should have administrator access to the system so they can view results, generate reports, and email reports to users.

Desired requirements include having the system correctly categorize results, generate reports, and email the reports to users and The Sum. Desired requirements also include implementing the paid and organizational versions of the PDA, moving the system to the cloud for scalability, and enhancing the systems administrator experience for The Sum.

Optional requirements include integrating the consultant scheduling system with the PDA system, supporting mobile devices, supporting changing the PDA questions, letting The Sum give consultants permissions to view specific user's results within the system, and allows organizations to view the results for their members who have taken the PDA.

STS Topic: Analyzing the Promise of Software Tools and Artificial Intelligence in Decreasing Bias and Establishing an Inclusive Work Environment

Although the nature of organizations operating in the D&I industry has shifted substantially in the past half-century (Anand & Winters, 2008) the methods and tools they deploy have remained nearly the same. That status quo, however, was disrupted when various diversity-related controversies, including the #MeToo movement, hit the tech sector (Reader, 2016). This led major Silicon Valley companies to reevaluate and reinvest in their diversity initiatives, which, paired with the recent advancements in machine learning algorithms, laid down and strengthened the foundations for the D&I Tech industry (Garr & Jackson, 2019). Currently, this industry is evaluated to be worth more than $100 million (p. 16) and is comprised of around 200 companies, 80% of which are less than 10 years old (p. 13). Some of the better-known vendors include Paradigm, HireVue, Pymetrics, Textio, which boast client lists that feature Twitter, Slack, Spotify, Mastercard, etc. The solutions - which include tools such as communication apps, analytics software, mentor matching algorithms - have the potential to vastly improve the hiring, retention and career advancement of employees from diverse

backgrounds (see figure 2). To limit scope, the STS research will focus on the subsection of the industry that specializes in AI enhanced recruitment. Not only is this the area most (43%) D&I tech companies offer services in (p. 22), it also happens to be the one that has received the greatest critical attention.

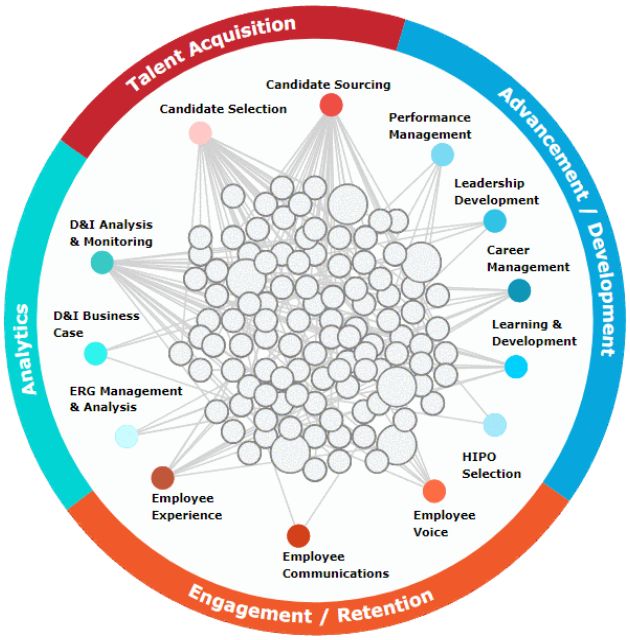

Figure 3: The categories of solutions provided by the D&I tech industry. Most vendors provide services across multiple categories (Garr & Jackson, 2019, p. 21)

Since hiring is the first obstacle to ensuring a diverse workforce, the heightened attention is palpable, especially when viewed under the light of statistical evidence. (To give just one example, in a 2017 study, Quillian, Pager, Hexel, & Midtbøen found "no change in the levels of discrimination against African Americans since 1989"). Moreover, a literature review on the use of AI in decision-making processes as varied as loan approval, bail assessments, stock trading, and candidate selection has shown that AIs deliver "more-efficient and more-equitable outcomes", even when the results are not attuned for bias in data (Miller, 2018). In one study, the candidates picked by AIs were “14% more likely to pass interviews and receive a job offer” and “18% more likely to accept job offers when extended” (Cowgill, 2019, p. 1). Compared to their

human counterparts, they seem to be far more consistent, even if not completely impartial, and can often help unearth useful patterns not obvious to human observers.

Critics, however, warn against such open-armed embrace, citing ever-present algorithmic bias stemming from skewed data, inconsistent selection of preferred attributes, and in general, the black-box nature of AIs (Bogen & Rieke, 2018; Raghavan, Barocas, Kleinberg, & Levy, 2019). Legal scholars have raised concerns about what these new systems may entail in interpreting anti-discrimination laws established in the 80s, when AIs were more fiction than science (Bent, 2019; Kleinberg, Ludwig, Mullainathan, & Sunstein, 2019). There are even real-world events to back up such unease; for instance, in 2018, Amazon scrapped an AI system that they claimed would be the "holy grail of hiring", when it started showing preferential treatment towards men (Dastin, 2018). And in 2019, the D&I tech vendor HireVue was sued by an employment rights group that claimed the results its algorithm generated were "biased, unprovable and not replicable" (Harwell, 2019).

These wide-ranging opinions about the role AIs can and cannot play in regards to bias-free hiring will be the focus of the STS research. Through an actor-network approach, the research will try to gauge the cultural, technological and organizational factors that are leading to the current state of "shared uncertainty" (Macospol, 2007). To structure the process, it will employ the observational lenses Venturini (2010) prescribed for exploring cartographies of controversy. As such, it will start off by looking into the "statements" made for and against the use of AI in recruitment, and then survey the "literatures", produced by experts in data science, organizational psychology, HR management, and law, from which the statements emerge. Next, it will look into the "actors" embedded within those literatures, whether they be vendors or their clients, data scientists or their algorithms, Title VII or its opponents and defendants. From there

it will attempt to unearth the inter-actor connections that form the backbone of the various "networks". After that, it will look into the ideologies or "cosmoses" that guide the actors' actions, such as thoughts on affirmative action or views on automation replacing human labor. Finally, in line with the concept of "cosmopolitics", it will examine and contrast the various pathways, each with its unique set of compromises, that may eventually lead to a "common world" (p. 268).

Conclusion

The deliverable of the technical project should be a safe, user-friendly, and maintainable website where users can take the PDA, and administrators can utilize the results to improve the underlying model. The STS project, on the other hand, should yield a better understanding of the controversy that surrounds new software systems pursuing the same goal of reducing institutional human biases, but within a much more complex sociotechnical system. If these objectives can be successfully met, the diversity industry would walk away with a better tool to raise awareness among people about the role they currently play, and can play in creating a more just, equitable world, while organizations in every industry will gain a framework to better understand what other such tools they should or should not include in their toolbox, as they pursue the long unfulfilled goal of a diverse and inclusive workplace.

References

Anand, R., & Winters, M.-F. (2008). A retrospective view of corporate diversity training from 1964 to the present. *Academy of Management Learning & Education, 7*(3), 356–372.

Bent, Jason R., Is algorithmic affirmative action legal? (April, 2019). *Georgetown Law Journal* [Forthcoming].

Bobo, L., Kluegel, J. R., & Smith, R. A. (1997). Laissez-faire racism: The crystallization of a kinder, gentler, antiblack ideology. In *Racial attitudes in the 1990s: Continuity and change*. Westport; London: Praeger.

Bogen, M., & Rieke, A. (2018). *Help wanted—An exploration of hiring algorithms, equity, and bias*. [Report] Retrieved from Upturn website: https://www.upturn.org/reports/2018/hiring-algorithms/

Cave, S., Craig, C., Dihal, K., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). *Portrayals and perceptions of AI and why they matter*. London: The Royal Society.

Cowgill, B. (2019). *Bias and productivity in humans and machines* [Working Paper] Retrieved from Upjohn Institute Working Paper website: https://research.upjohn.org/up_workingpapers/309/

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Faigman, D. L., Dasgupta, N., & Ridgeway, C. L. (2007). The matter of fit: The law of discrimination and the science of implicit bias symposium. *Hastings Law Journal, 59*(6), 1389–1434.

Garr, S. S., & Jackson, C. (2019). *Diversity & inclusion technology: The rise of a transformative market* [Report]. Retrieved from Mercer website: https://www.mercer.com/our-thinking/career/diversity-and-inclusion-technology.html

Harwell, D. (2019, November 6). Rights group files federal complaint against AI-hiring firm HireVue, citing ‘unfair and deceptive’ practices. *Washington Post*. Retrieved from https://www.washingtonpost.com/technology/2019/11/06/prominent-rights-group-files-federal-complaint-against-ai-hiring-firm-hirevue-citing-unfair-deceptive-practices/

Hebl, M., Ruggs, E., Martinez, L., Trump-Steele, R., & Nittrouer, C. (2016). Understanding and reducing interpersonal discrimination in the workplace. In T. D. Nelson (Ed.), *Handbook of Prejudice, Stereotyping, and Discrimination* (pp. 387-407). New York, NY, US: Psychology Press.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2019). Discrimination in the age of algorithms. *Journal of Legal Analysis*, *10*, 113–174.

Krysan, M., & Moberg, S. (2016, August 25). *Trends in racial attitudes.* University of Illinois Institute of Government and Public Affairs. Retrieved from http://igpa.uillinois.edu/programs/racial-attitudes

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*(2), 275-287.

Miller, A. P. (2018). Want less-biased decisions? Use algorithms. *Harvard Business Review Digital Articles*, 2–5.

Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, *3*(3), 314–338.

Project Implicit (n.d.). Retrieved December 13, 2019 from https://implicit.harvard.edu/implicit/blog.html

Quillian, L., Pager, D., Hexel, O., & Midtbøen, A. H. (2017). Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(41), 10870–10875.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2019). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *arXiv:1906.09208 [Cs]*. Retrieved from http://arxiv.org/abs/1906.09208

Reader, R. (2016, July 5). How silicon valley's diversity problem created a new industry. *Fast Company.* Retrieved October 31, 2019 from https://www.fastcompany.com/3061457/how-silicon-valleys-diversity-problem-created-a-new-industry

The Sum. (n.d.). Retrieved October 22, 2019, from https://www.thesum.org/