Harm and Data Science: Framing Negative Human Outcomes as a Data Science Problem

Katherine Weathington
katy.weathington@colorado.edu
Department of Information Science
University of Colorado Boulder
Boulder, Colorado, United States

Morgan Klaus Scheuerman
morgan.scheuerman@colorado.edu
Department of Information Science
University of Colorado Boulder
Boulder, Colorado, United States

ABSTRACT

After researching predictive policing practices, and with police violence an ever present horror, the first author found herself motivated to consider the real human harm that can result from data driven decisions. We examine how certain data science practices associated with predictive policing contribute to and are complicit with human suffering caused by police brutality. We then put forward an idea of data scientists internally motivated to center their work around minimizing negative human consequences rather than traditional goals of performance optimization or increasing profits. We encourage further research into the disciplinary norms which distance model builders from practical implementations of their work, and how academics can intervene.

INTRODUCTION

Imagine you are a data scientist, tasked with analyzing crime data to decide where patrols will be. Further imagine that an officer you have placed pulls the trigger, and ends a life. Had you not decided so, this officer would not have been there to kill someone. How would you feel? Would you be wracked with guilt and resign, or motivated to change your practices? Would you even make the connection between the product of your labor and such extreme violence? Unfortunately, we live in a reality where this hypothetical could be all too real.

Over the past several years, the first author has conducted research on the implementation and use of geospatial predictive policing tools. A large portion of her work has examined how police employees implemented geospatial analysis software and how the reports are then interpreted and converted into actual patrolling allocation decisions, which led to many interactions with crime analysts, the data scientists specifically tasked with analyzing crime data. In addition to exploring *how* these tools are being used by police, she questioned *why* these tools are being used. Understanding the motivations for predictive policing allows researchers to better understand the specific implementation practices chosen as well as provide feedback in a way that is more likely to be valued and implemented.

After the first author's various studies into predictive policing algorithms, and all the problems with them, the new wave of racist police violence inspired a new set of questions. The mental juxtaposition of those data scientists in crime labs and the armor clad riot police, both working for the same system, stuck in her head. How do these data scientists understand their place and connection with state sanctioned tools of violence? Do they recognize the role they might have played at all? Do they feel culpable or even guilty? Do they feel a motivation to make meaningful reform from within? You could also ask all of these questions of the data scientists at tech firms who built the predictive policing tools in the first place.

Questions such as these have haunted us for the past several months, and we have concentrated them into several questions for discussion and future research both within predictive policing and generalized to the field of data science as a whole. What are the real, human impacts of data science practices? Do data scientists see themselves as having a role in such impacts at all? If so, what do they feel about it? What can individual data scientists do to make meaningful changes? In particular, how can data scientists, who through their work reduce humans to mere data points, reconcile with the possibility of complicity in the destruction of human life? With industry and most of the public sector prioritizing performance optimization, we question whether it is possible for data scientists to make this change from the inside, or whether it is necessary to divest themselves from their corporate or governmental backers.

DATA-DRIVEN POLICING PRACTICES

Research has shown that officers who are aware of being in an algorithmically identified 'hotspot', that is, a geographically defined area with a relatively high rate of past or predicted crime, are more likely to take biased police action against racial minorities than those who are not in a hotspot [13]. Furthermore, it has been shown that there are significant racial biases in whom police choose to interact with and the extent of action taken in almost every aspect of policing [4–7, 10, 12, 14, 16, 18, 19, 21, 23–27]. In combination, this means that an officer who is placed in an algorithmically identified hotspot by data driven software are more likely to take drastic, potentially violent action against racial minorities. Were an algorithmically placed officer to take such a heightened action and cause serious, irreparable harm, then the data scientists who put them there would be a major link in the causal chain for that harm, creating an ethical quandry for the data scientist and possibly leading to intense guilt.

There are several further factors in the deployment of predictive policing tools which compound the ethical qualms of such scenarios. Firstly, many predictive policing and data driven criminal justice tools are hidden behind black boxes, making it incredibly difficult to audit tools being used and verify any level of impartiality and prevent contesting decisions [2, 9]. Second, many crime lab workers lack specific training with and an intuitive understanding of the tools they implement [28]. Both of these highlight bad data science practices which absolutely could lead to biased arrests, police brutality, and even state sanctioned killings.

Often, the stated motive for using predictive policing tools was exclusively to increase officer efficiency [11]. However, there was a significant lack of attention to reducing bias in policing actions. Such goals demonstrate a failure to center human outcomes, instead focusing on saving money through promoting a purely technical concept of performance. Such a paradigm will clearly create ethical pitfalls; that is, some scenario where the immediate outcome is not harmful, but creates a scenario that can easily become incredibly harmful with the right circumstances.

Such ethical pitfalls are not limited to predictive policing itself. Recently, a data driven advertising service posted many Los Angeles Police Department recruitment ads on the alt-right website Breitbart [22], which could lead to a higher number of extremely conservative, arguably fascist readers joining the police, and bring their own, often racially biased, political ideologies to what should essentially be an impartial and objective position. Even if the ad placements were based on pre-existing biases of American police, the data driven advertising creates a feedback loop of self-reinforcing biases. We can see the effects of this radicalizing of law enforcement in the Black Lives Matters (BLM) protests, where heavily armed right wing militia groups are supported and provisioned by the same police who teargas BLM protesters, even to the point of letting a right wing militia member kill two protesters and walk away [15]. Perhaps if recruitment advertisements had been on a more left wing web site, the bias would have been lessened.

CREATING HUMAN-FIRST SYSTEMS

While much has been said about biased outcomes of data science practices, it has often been in an abstract vacuum rather than grounded in the actual negative consequences of bias [8]. This allows data scientists to be able to justify almost any tool with some arbitrary accuracy rating, which reflects merely a mathematical concept of correctness and allows data scientists to distance themselves from the actual fallout of their work. It becomes a game of what one can do rather than what one should do, and is how we end up with dubious projects such as algorithmic ethnicity classifiers [17, 20], which could easily be used as racist profiling tools.

A wide range of significant biases exist in the tools and practices created by data scientists. While many academics call for more equitable and unbiased systems, such a sentiment is rarely heard from non-academic working data scientists. And as data driven tools are implemented in increasingly impactful contexts [1], data scientists have an increasing ethical duty to build systems which minimize the harm they cause, while centering that work around specific cases of human suffering rather than optimizing performance. Though this *can* mean the same thing, such as with medical diagnostic tools, often one comes at the cost of the other.

Such a shift in motivation would encourage data scientists to know where data science does not belong at all. Not every situation calls for a big data machine learning tool [3], yet building such a tool for every possible use leads to a wide variety of over-promising, under-performing tools which are never truly checked for the implications of their use. However, efforts limiting the availability of data science tools run opposed to the need for open, transparent tools. And thus such an attitude change must come from within the field itself and should be fostered by better ethical education practices. This raises a question of how academics can reduce the proliferation of garbage data science without hiding behind the blackboxes that cause us so many headache and without gatekeeping data science.

To allow for such a change requires acknowledging and changing one not-so-implicit bias of the data science field as a whole. As most tools are built at the behest of corporations, whether in house or at a dedicated software firm, these tools overwhelmingly serve to protect capital, increase revenue, and essentially reduce their human subjects to data points of their economic value. Thus we can visualize the data scientist pulled between two realities: the need to satisfy their employers, and the desire to follow their conscience. The academic community must consider this conflict when pushing for theory based solutions and devising new tools. Without acknowledging that first reality, academics are creating solutions only for academia, and prevent solutions from adequately translating to industry or the public sector. Without just leaving a company, as is often called for, what can a data scientist without a background in critical algorithm analysis realistically do while keeping their job? We suggest further discussion and research into empowering data scientists to reduce harm centered around real, grounded understandings of human harm.

CONCLUSION

Grounded in a background in predictive policing, and in light of the widespread racial police brutality, we see how data science can contribute to human suffering and even death. We see how even in other contexts, the data science field has incredible power to destroy lives. Thus, we call for a critical examination of the systemic biases and practices in the field of data science which allow them to make systems which harm, and call for a paradigm shift towards building systems with an explicit of goal centering and empowering people. We ask what practicing data scientists, grounded in the technical aspects of data science rather than theory and ethics, can do? What changes can be made from within? What tools do they need to make these changes? What can academia do to facilitate real data scientists to make meaningful and ethical changes themselves?

REFERENCES

[1] 2015. Privacy expert argues "algorithmic transparency" is crucial for online freedoms at UNESCO knowledge café. (Dec 2015). https://en.unesco.org/news/privacy-expert-argues-algorithmic-transparency-crucial-online-freedoms-unesco-knowledge-cafe

[2] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. Machine Bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=sYBNO6t1202JOb6ILFkA_eTWzPmpol3N

[3] Eric P.S. Baumer and M. Six Silberman. 2011. When the Implication is Not to Design (Technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2271–2274. https://doi.org/10.1145/1978942.1979275

[4] Frank R. Baumgartner, Derek A. Epp, Kelsey Shoub, and Bayard Love. 2017. Targeting young men of color for search and arrest during traffic stops: evidence from North Carolina, 2002–2013. *Politics, Groups, and Identities* 5, 1 (2017), 107–131. https://doi.org/10.1080/21565503.2016.1160413

[5] Katherine Beckett, Kris Nyrop, and Lori Pfingst. 2006. Race, drugs, and policing: Understanding disparities in drug delivery arrests. *Criminology* 44, 1 (2006), 105–137.

[6] Benjamin Bowling. 2018. Pulled over: how police stops define race and citizenship. *Policing and Society* 28, 5 (2018), 626–629. https://doi.org/10.1080/10439463.2018.1473978

[7] James W Buehler. 2017. Racial/ethnic disparities in the use of lethal force by US police, 2010–2014. *American journal of public health* 107, 2 (2017), 295–297.

[8] David Danks and Alex John London. 2017. Algorithmic Bias in Autonomous Systems. International Joint Conferences on Artificial Intelligence Organization, 4691–4697. https://doi.org/10.24963/ijcai.2017/654

[9] Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science Advances* 4, 1 (Jan 2018). https://doi.org/10.1126/sciadv.aao5580

[10] Ronnie A Dunn. 2009. Measuring racial disparities in traffic ticketing within large urban jurisdictions. *Public Performance & Management Review* 32, 4 (2009), 537–561.

[11] Steven J. D'Amico, Shoou-Jiun Wang, Rajan Batta, and Christopher M. Rump. 2002. A simulated annealing approach to police district design. *Computers Operations Research* 29, 6 (2002), 667–684. https://doi.org/10.1016/S0305-0548(01)00056-9

[12] Jeffrey Fagan, Anthony A Braga, Rod K Brunson, and April Pattavina. 2016. Stops and stares: Street stops, surveillance, and race in the new policing. *Fordham Urb. LJ* 43 (2016), 539.

[13] Andrew Guthrie Ferguson. 2012. Predictive policing and reasonable suspicion. *Emory LJ* 62 (2012), 259.

[14] Sharad Goel, Justin M. Rao, and Ravi Shroff. 2016. Precinct or prejudice? Understanding racial disparities in New York City's stop-and-frisk policy. *The Annals of Applied Statistics* 10, 1 (March 2016), 365–394. https://doi.org/10.1214/15-AOAS897

[15] Isiah Holmes. 2020. Disparate treatment of militia, protesters in Kenosha. https://wisconsinexaminer.com/2020/08/27/disparate-treatment-of-militia-protesters-in-kenosha/

[16] William C. Horrace and Shawn M. Rohlin. 2016. How Dark Is Dark? Bright Lights, Big City, Racial Profiling. *The Review of Economics and Statistics* 98, 2 (2016), 226–232. https://doi.org/10.1162/REST_a_00543

[17] S. Hosoi, E. Takikawa, and M. Kawade. 2004. Ethnicity estimation with facial images. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.* 195–200.

[18] Tammy Rinehart Kochel, David B Wilson, and Stephen D Mastrofski. 2011. Effect of Suspect Race on Officers' Arrest Decisions. *Criminology* 49, 2 (2011), 473–512.

[19] Joscha Legewie. 2016. Racial Profiling and Use of Force in Police Stops: How Local Events Trigger Periods of Increased Discrimination. *Amer. J. Sociology* 122 (09 2016), 379–424. https://doi.org/10.1086/687518

[20] Xiaoguang Lu, Anil K Jain, et al. 2004. Ethnicity identification from face images. In *Proceedings of SPIE*, Vol. 5404. 114–123.

[21] Mona Lynch, Marisa Omori, Aaron Roussell, and Matthew Valasik. 2013. Policing the 'progressive'city: The racialized geography of drug law enforcement. *Theoretical criminology* 17, 3 (2013), 335–357.

[22] Sanya Mansoor. 2019. *A Recruitment Ad for the Los Angeles Police Department Showed Up on Breitbart. Now, the LAPD Is Investigating.* https://time.com/5688815/los-angeles-police-department-ad-breitbart/

[23] Justin Nix, Bradley A Campbell, Edward H Byers, and Geoffrey P Alpert. 2017. A bird's eye view of civilians killed by police in 2015: Further evidence of implicit bias. *Criminology & Public Policy* 16, 1 (2017), 309–340.

[24] Emma Pierson, Camelia Simoiu, Jan Overgoor, Sam Corbett-Davies, Daniel Jenson, Amy Shoemaker, Vignesh Ramachandran, Phoebe Barghouty, Cheryl Phillips, Ravi Shroff, and et al. 2020. A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour* 4, 77 (Jul 2020), 736–745. https://doi.org/10.1038/s41562-020-0858-1

[25] E Ashby Plant and B Michelle Peruche. 2005. The consequences of race for police officers' responses to criminal suspects. *Psychological Science* 16, 3 (2005), 180–183.

[26] Forrest Stuart. 2016. *Down, Out, and Under Arrest.* https://www.press.uchicago.edu/ucp/books/book/chicago/D/bo23530208.html

[27] Aatricia Warren, Donald Tomaskovic-Devey, William Smith, Matthew Zingraff, and Marcinda Mason. 2006. DRIVING WHILE BLACK: BIAS PROCESSES AND RACIAL DISPARITY IN POLICE STOPS*. *Criminology* 44, 3 (2006), 709–738. https://doi.org/10.1111/j.1745-9125.2006.00061.x

[28] Katherine Weathington. 2020. A Critical Study of Geospatial Algorithm Use in Crime Analysis and Predictive Policing. (Apr 2020). https://epublications.marquette.edu/theses_open/588