Translation Tutorial: Toward a Theory of Race for Fairness in Machine Learning

Emanuel D. Moss
CUNY Graduate Center
Data & Society
New York, New York
emoss@gradcenter.cuny.edu

ABSTRACT

While computer scientists working on questions of fairness have diligently produced algorithmic approaches that seek to minimize disparate impacts across racial categories, the concept of race itself remains either unexamined, or constrained by definitions arising in legal and policy domains. While this may be appropriate for some applications, it is not altogether obvious that the FAT community benefits from refraining from developing a theory of race to guide its own practices. This tutorial will translate concepts from critical race theory and social scientific discourses into concepts legible to a community of machine learning practitioners through a discussion of these theories and small-group activities that illustrate the salience of these theories for problems of fairness in machine learning.

1 TEAM

Emanuel Moss - Doctoral Candidate in Cultural Anthropology at CUNY Graduate Center and Research Analyst at Data & Society.

2 DESCRIPTION

While computer scientists working on questions of fairness have diligently produced algorithmic approaches that seek to minimize disparate impacts across classes (particularly across race and gender classes), the entire FAT community seems to be operating entirely without a theory of race (or gender, or ethnicity, or sexuality, or etc.). In consequence, most of the efforts around fairness take the social categories at hand as given, and make little consideration of what the various categories they work with represent, how they have come to be constructed as such, or that the concepts of race and gender are themselves social constructs. In the absence of a theory of race, approaches that try to generalize across different data sets are often inadequate, and sometimes lead to the problematic use of other biological features as proxies for race in some analyses, as

well as the resurgence of approaches that look a lot like a new form of physiognomy [1].

This tutorial will center on a discussion of possibilities for what a theory of race could look like for machine learning. It will do so by opening with a presentation of several approaches to understanding race and the history of racialization in the United States, how that history has been inseparable from data collection practices from the 18th Century to the present day, and how quantificatory practices, scientific understandings, and sociopolitical movements have all contributed to the statistical and computational practices that undergird machine learning.

A key focus of this presentation will be the history of social statistics as a system of classification, and how this history has been replicated by data science and machine learning [3]. It will also address race as a social construct that is made real through a host of practices, including its use as a classificatory system [6] [2]. The presentation will also address the implications of machine learning are for understandings of race as inextricably bound up with the body, presentations of self, and genetic ancestry [4][5][7].

The tutorial itself will begin with a seminar-style discussion of the importance of a theory of race. The rest of the tutorial will center on an exercise in which participants will be asked to work through a scenario relevant to machine learning in which race functions in a meaningful way. The scenario will be presented for discussion by small groups of participants, who will then present their discussion to the full group.

- Scenario : The U.S. Census has changed its methodology for collecting data on race from one census to the next, shifting the labels applied to various racial groups, as well as the number of racial divisions. How would you, as a machine learning researcher, address this shifts in classificatory practices? What do these shifts tell you about the data itself? What do these shifts tell you about the underlying phenomena represented by census data?

3 TIMELINE

Total Time: 45 minutes

- 20 minutes: Opening panel discussion of history of race and classification systems.
- 12 minutes: Scenario 1 small-group work.
- 12 minutes: Scenario 1 full group discussion.

4 A/V NEEDS

None

Emanuel D. Moss

REFERENCES

[1] Blaise Agüera y Arcas, Margaret Mitchell, and Alexander Todorov. 2017. Physiognomy's New Clothes. (2017).

[2] Kwame Anthony Appiah and Amy Gutmann. 1996. *Color Conscious: The Political Morality of Race*. Princeton University Press, Princeton, NJ.

[3] Geoffrey C. Bowker and S. L. Star. 2000. *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.

[4] Lundy Braun, Anne Fausto-Sterling, Duana Fullwiley, Evelynn M Hammonds, Alondra Nelson, William Quivers, Susan M Reverby, and Alexandra E Shields. 2007. Racial Categories in Medical Practice: How Useful Are They? *PLOS Medicine* 4, 9 (09 2007), 1–6. https://doi.org/10.1371/journal.pmed.0040271

[5] Duana Fullwiley. 2007. *The Molecularization of Race: Institutionalizing Human Difference in Pharmacogenetics Practice*. Vol. 16. 1–30 pages. https://doi.org/10.1080/09505430601180847

[6] F. V Harrison. 1995. The Persistent Power of 'Race' in the Cultural and Political Economy of Racism. *Annual Review of Anthropology* 24, 1 (1995), 47–74. https://doi.org/10.1146/annurev.anthro.24.1.47

[7] John Hartigan Jr. 2008. Is Race Still Socially Constructed? The Recent Controversy over Race and Medical Genetics. *Science as Culture* 17, 2 (2008), 163–193. https://doi.org/10.1080/09505430802062943