INCLUSIVE & PARTICIPATORY DESIGN IN ML

NOV 11, 2020

IACS CAPSTONE

SPEAKERS:

WEIWEI PAN
RESEARCH ASSOCIATE

MATTHEW FINNEY
GRADUATE STUDENT ADVISORY COMMITEE

SCARLETT GONG
GRADUATE STUDENT ADVISORY COMMITEE

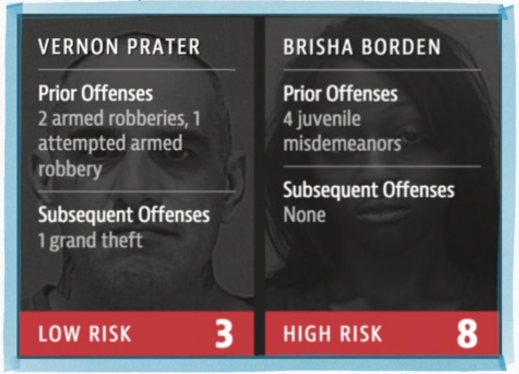
PAUL TEMBO
GRADUATE STUDENT ADVISORY COMMITEE

EXAMPLES OF DATA SCIENCE FAILS:

PREDICTIVE POLICING

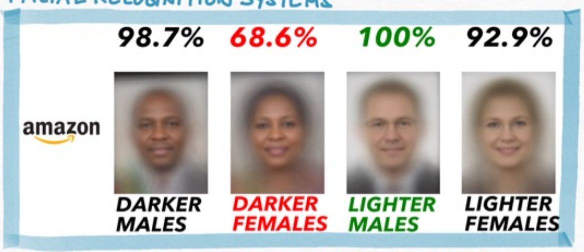
PREDICTIVE POLICING ALGORITHMS ARE RACIST. THEY NEED TO BE DISMANTLED, MIT TECHNOLOGY REVIEW 2020

PRE-TRIAL, PAROLE RISK ASSESMENT



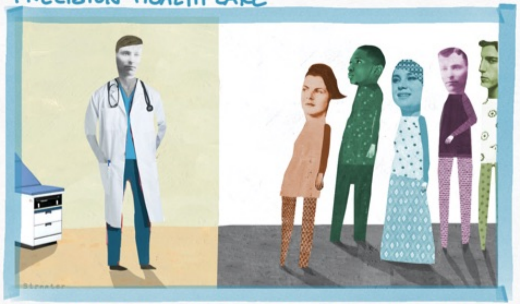
INJUSTICE EX-MACHINA: PREDICTIVE ALGORITHM IN CRIMINAL JUSTICE, UCLA LAW REVIEW, 2019

FACIAL RECOGNITION SYSTEMS



FACIAL RECOGNITION IS ACCURATE, IF YOU'RE A WHITE GUY, NEW YORK TIMES, 2018

PRECISION HEALTH CARE



SEX AND GENDER DIFFERENCES AND BIASES IN ARTIFICIAL INTELLIGENCE FOR BIOMEDICINE AND HEALTH CARE, NATURE, 2020

A LONG HISTORY OF DESIGN FAILS:

PROBLEMS WITH DATA SCIENCE IS NOT AN EXCEPTION, NOT A GLITCH, IT'S PART OF A LONG HISTORY OF SYSTEMATIC DESIGN & DEPLOYMENT FAILS OF NON-DIGITAL TECH.

RACE AFTER TECHNOLOGY, RUHA BENJAMIN

- PHOTOGRAPHY IS A TECH OF SUBJECTIVE DECISIONS
- CALIBRATION OF COLOR PHOTOS USES A "SHIRLEY" CARD REFERENCE WITH WHITE WOMAN
- RENDER DARKER SKIN FEATURES BLURRY
- 1960'S COMPLAINTS FROM BLACK PARENTS ABOUT POOR QUALITY SCHOOL PHOTOS TO KODAK
- 1960'S, 70'S COMPLAINTS FROM FURNITURE & CHOCOLATE COMPANIES FINALLY CHANGED COLOR CALIBRATION PROCEDURES

INVISIBLE WOMEN, CAROLINE CRIADO PEREZ

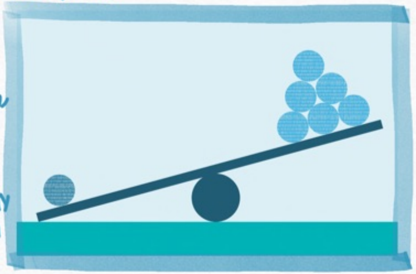
- OVER 80% OF GLOBAL POPULATION STILL USE TRADITIONAL STOVES & BIOMASS FUEL.
- TRADITIONAL STOVES GIVE OFF TOXIC FUMES, EQUIVALENT TO >100 CIGARETTES A DAY IN UNVENTILATED ROOM.
- TRADITIONAL STOVES ARE INEFFICIENT, WOMEN COOK FOR 3-7 HOURS A DAY.
- EFFORTS TO DESIGN & POPULARIZE EFFICIENT, CLEAN STOVES STARTED IN 1950'S, LARGELY RESULTED IN FAILURES

WHY IS THIS HAPPENING?

BIAS IN TECH IS NOT AN ALGORITHMIC, MATH PROBLEM; IT IS A PEOPLE PROBLEM, IT IS ABOUT NEGLECTING THE HUMAN DIMENSION AND INEXTRICABLY LINKED WITH DIVERSITY & REPRESENTATION.

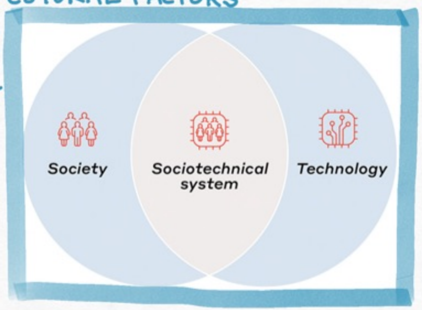
DATA GAP

- LACK OF SEX AGGREGATED DATA IN ECON, URBAN PLANNING, MEDICINE, AGRICULTURE, ETC.
- LACK OF DIVERSITY IN ML TRAINING DATA

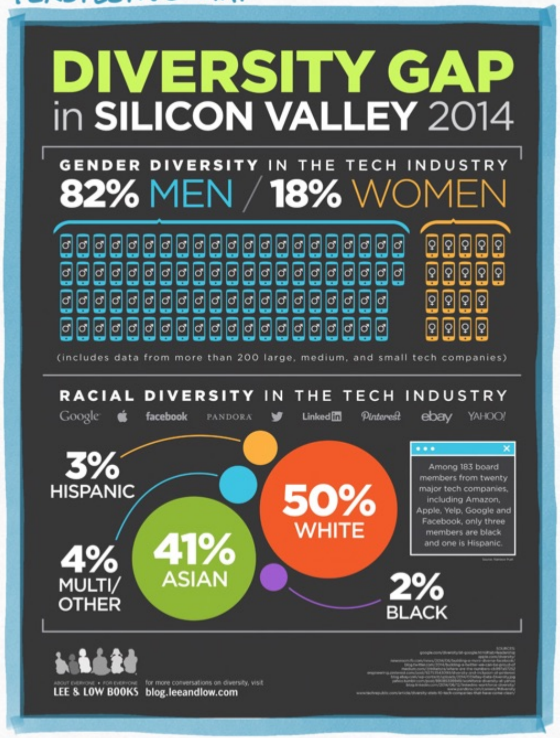


NEGLECTING SOCIAL, POLITICAL CUTURAL FACTORS

- NO TECH STANDS ALONE WITHOUT SOCIAL, CULTURAL POLITICAL, ECON HISTORICAL CONTEXT
- NO PROBLEM IS SOLUABLE BY TECH ALONE



PERSPECTIVE GAP



- DESIGN CHOICES ARE NOT NEUTRAL NOR OBJECTIVE BUT ENCODE VALUES AND PERSPECTIVES
- PRIORITIZE TECHNICAL PARAMETERS OVER USER NEEDS
- COMMUNITIES THAT ARE IMPACTED BY TECH ARE NOT REPRESENTED AS DESIGNERS, POLICY MAKERS, ENGINEERS

RACE AFTER TECHNOLOGY, RUHA BENJAMINE

INVISIBLE WOMEN, CAROLINE CRIADO PEREZ

DOES FAIRNESS RESEARCH FIX IT?

FAIRNESS DEFINED MATHEMATICALLY ABSENT REAL-LIFE CONTEXT AND ENGAGEMENT OF STAKEHOLDERS CAN REINFORCE EXISTING BIASES AND INJUSTICES: FAIRNESS WASHING

BEYOND BIAS: REIMAGINING THE TERMS OF ETHICAL AI IN CRIMINAL LAW

- FAIRNESS RESEARCH HAPPENING IN ACADEMIA, MAJOR TECH CORPS, FAR FROM AFFECTED COMMUNITIES.
- FOCUS ON TECHNICAL DEFINITION OF PROBLEM, E.G. DATA IMBALANCE, SUBGROUP DISPARITY, AND TECHNICAL SOLUTIONS, NOT ON SOCIO-TECH SYSTEM:
  - ↳ DEFINITION OF CRIME: STREET V.S. WHITE COLLAR
  - ↳ CONFLATION OF ARREST & DANGER
  - ↳ RACIAL DISPARITIES IN POLICING, ARREST, CHARGING, LEGAL REPRESENTATION SENTENCING PRACTICES

ETHICAL LIMITS OF ALGORITHMIC FAIRNESS SOLUTIONS IN HEALTH CARE MACHINE LEARNING

- FAIRNESS METRICS DO NOT TAKE INTO ACCOUNT COMPLEX CAUSAL RELATIONSHIPS BTW BIOLOGICAL ENVIROMENTAL & SOCIAL FACTORS OF MEDICAL CONDITIONS.
- DIFFERENCE DOES NOT ENTAIL INEQUITY IN MEDICAL TREATMENT. BLINDING DECISIONS TO PROTECTED ATTRIBUTES NOT ALWAYS GOOD
- FAIRNESS METRICS CAN FORCE ALIGNMENT BTW HETEROGENEOUS GROUPS
- FAIR DOES NOT MEAN HARMLESS, WHICH IS THE FIRST GOAL OF MEDICINE

BETTER DESIGN THROUGH PARTICIPATION:

MACHINE LEARNING IS MAKING A MOVEMENT AWAY FROM TRADITIONAL DESIGN FRAMEWORKS.

QUESTIONS FOR ETHICAL ML

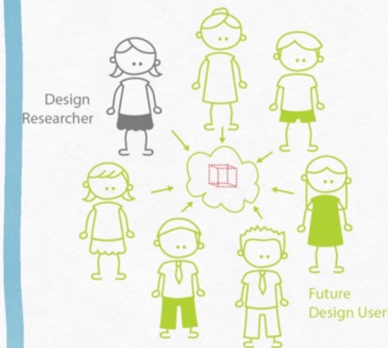
I. IDENTIFY THE STAKEHOLDERS
- WHO ARE THE USERS?
- WHO ARE THE AFFECTED COMMUNITIES?

II. WHAT TYPES OF HARM CAN YOUR TECH DO?
- WHAT KIND OF HARM CAN TECH FAILURES CAUSE
- WHAT KIND OF HARM CAN THE SOCIO-TECHNICAL SYSTEM CAUSE

III. WHAT TYPES OF GOOD CAN YOUR TECH DO?
- WHAT KINDS OF NEEDS DO YOUR USERS HAVE
- WHAT KINDS OF CONSTRAINTS DO YOUR USERS HAVE

IV. WHAT ARE YOUR OWN ETHICAL PROFESSIONAL RESPONSIBILITIES?
- WHAT IS ENGINEERING ETHICS?
- DO ENGINEERS NEED ETHICS?

PARTICIPATORY FRAMEWORKS FOR MACHINE LEARNING



- PARTICIPATORY DESIGN IS DESIGN THAT ACTIVELY INCORPORATE FEED BACK FROM STAKE-HOLDERS
- CO-DESIGN IS A COLLABORATIVE DESIGN PROCESS BTW DESIGNERS & STAKE-HOLDERS
- PARTICIPATORY ACTION RESEARCH IS COMMUNITY EMBEDDED RESEARCH THAT PARTNER WITH AFFECTED COMMUNITIES TO ACHIEVE SOCIAL CHANGE

LESSONS FROM PARTICIPATORY MACHINE LEARNING

1. DON'T MAKE SYSTEMS MORE FAIR AT THE EXPENSE OF MAKING THEM MORE JUST. GO BEYOND TECHNICAL FRAMINGS OF HARM, E.G. ALGORITHMIC BIAS, DATA IMBALANCE.
2. COMMUNITIES DERIVE MOST VALUES FROM LOCALIZED SOLUTIONS RATHER THAN SCALABLE, GENERALIZABLE SOLUTIONS.
3. MANY MEANINGFUL INTERVENTIONS TOWARDS MORE EQUITABLE SYSTEMS ARE NON-TECHNICAL.

TOWARDS SITUATED INTERVENTIONS FOR ALGORITHMIC EQUITY, KATELL ET AL
FAT* 2020

BEWARE OF PARTICIPATION-WASHING:

COMMUNITY CENTERED VS AT-SCALE

- MONETARY INCENTIVE IS FOR SOLUTIONS THAT SCALE & GENERALIZE
- FUNDS MUST BE SET-ASIDE TO LOCALLY UPDATE OR ADAPT MODEL AFTER SCALING

PARTICIPATION MUST BE LONG TERMED AND GENUINE

- PERFORMATIVE PARTICIPATION: EXPERTS DO NOT KNOW HOW TO DESIGN USEFUL PARTICIPATORY FRAMEWORKS & ENGAGE WITH STAKE HOLDERS
- USEFUL PARTICIPATORY DESIGN IS NOT WITHOUT FRICTION; REQUIRES CONSTANT UPDATING AND MAINTAINANCE

YOUR ETHICAL OBLIGATIONS AS A DATA SCIENTIST:

WHY SHOULD YOU AS AN INDIVIDUAL DATA SCIENTIST HAVE TO ENGAGE WITH THESE ISSUES?

NSPE CODE OF ETHICS

National Society of Professional Engineers Code of Ethics for Engineers

Preamble: Engineering is an important and learned profession. As members of this profession, engineers are expected to exhibit the highest standards of honesty and integrity. Engineering has a direct and vital impact on the quality of life for all people. Accordingly, the services provided by engineers require honesty, impartiality, fairness and equity, and must be dedicated to the protection of the public health, safety, and welfare. Engineers must perform under a standard of professional behavior that requires adherence to the highest principles of ethical conduct.

Fundamental Canons

Engineers, in the fulfillment of their professional duties, shall:

- Hold paramount the safety, health and welfare of the public.
- Perform services only in areas of their competence.
- Issue public statements only in an objective and truthful manner.
- Act for each employer or client as faithful agents or trustees.
- Avoid deceptive acts.
- Conduct themselves honorably, responsibly, ethically, and lawfully so as to enhance the honor, reputation, and usefulness of the profession

ASCE CODE OF ETHICS

1. SOCIETY

Engineers:

a. first and foremost, protect the health, safety, and welfare of the public;

b. enhance the quality of life for humanity;

c. express professional opinions truthfully and only when founded on adequate knowledge and honest conviction;

d. have zero tolerance for bribery, fraud, and corruption in all forms, and report violations to the proper authorities;

e. endeavor to be of service in civic affairs;

f. treat all persons with respect, dignity, and fairness, and reject all forms of discrimination and harassment;

g. recognize the diverse historical, social, and cultural needs of the community, and incorporate these considerations in their work;

h. consider the capabilities, limitations, and implications of current and emerging technologies when part of their work; and

READING LIST:

1. Angwin, Julia, et al., Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks, *ProPublica* (2016)
2. Barabas, Chelsea. "Beyond Bias: Re-Imagining the Terms of 'Ethical AI'in Criminal Law." *Chelsea Barabas, Beyond Bias: Re-imagining the Terms of" Ethical AI" in Criminal Law* 12 (2019).
3. Benjamin, Ruha. "Race after technology: Abolitionist tools for the new jim code." *Social Forces* (2019).
4. Black, Emily, and Michael A. Madaio. "A Call for Universities to Develop Requirements for Community Engagement in AI Research."
5. Brown, Anna, et al. "Toward algorithmic accountability in public services: A qualitative study of affected community perspectives on algorithmic decision-making in child welfare services." *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019.
6. Criado-Perez, Caroline. Invisible Women: Data Bias in a World Designed for Men. New York: Abrams Press, 2019.
7. Hill, Kashmir. Wrongfully Accused by an Algorithm, *The New York Times* (2020)
8. Katell, Michael, et al. "Toward situated interventions for algorithmic equity: lessons from the field." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020.
9. Lohr, Steve. Facial Recognition Is Accurate, if You're a White Guy, *The New York Times* (2018).
10. McCradden, Melissa D., et al. "Ethical limitations of algorithmic fairness solutions in health care machine learning." *The Lancet Digital Health* 2.5 (2020): e221-e223.
11. Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366.6464 (2019): 447-453.
12. Sloane, Mona, et al. "Participation is not a Design Fix for Machine Learning." *arXiv preprint arXiv:2007.02423* (2020).
13. "Predictive Policing Algorithms Are Racist. They Need to Be Dismantled." MIT Technology Review. 2020.
14. Green, Ben. "Data science as political action: grounding data science in a politics of justice." Available at SSRN 3658431 (2020).
15. Chen, Irene, Fredrik D. Johansson, and David Sontag. "Why is my classifier discriminatory?." Advances in Neural Information Processing Systems. 2018.