

# CS 4001: Computing, Society & Professionalism

Munmun De Choudhury | Assistant Professor | School of Interactive Computing

## Week 13: Algorithmic Bias

April 9, 2018

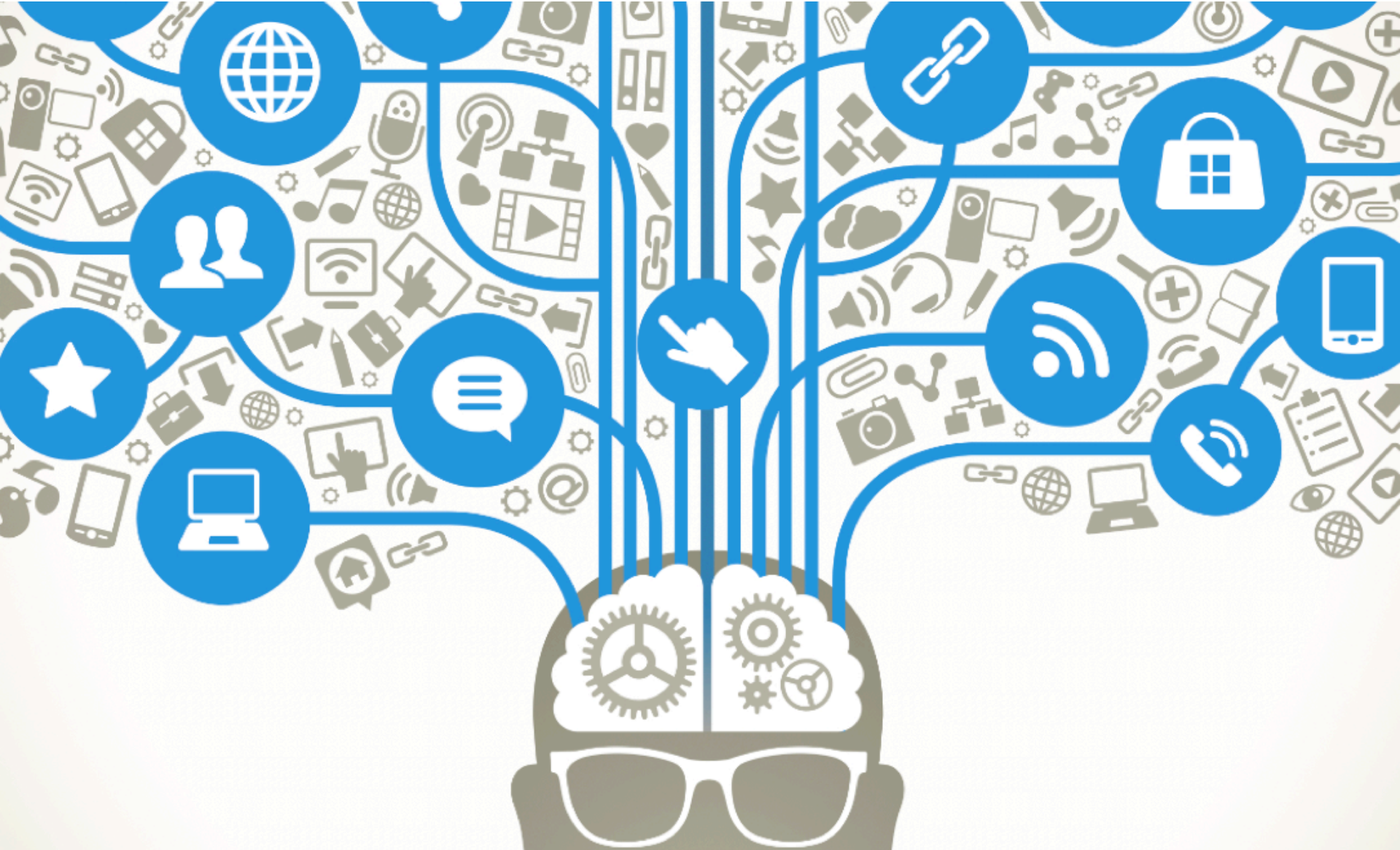
# Final Exam

- Take-home
  - Open book and internet, so relatively harder than midterm
- Example exercises and review material will be released after the last class
  - Similar in flavor to the class exercises





# Ethics of Algorithms



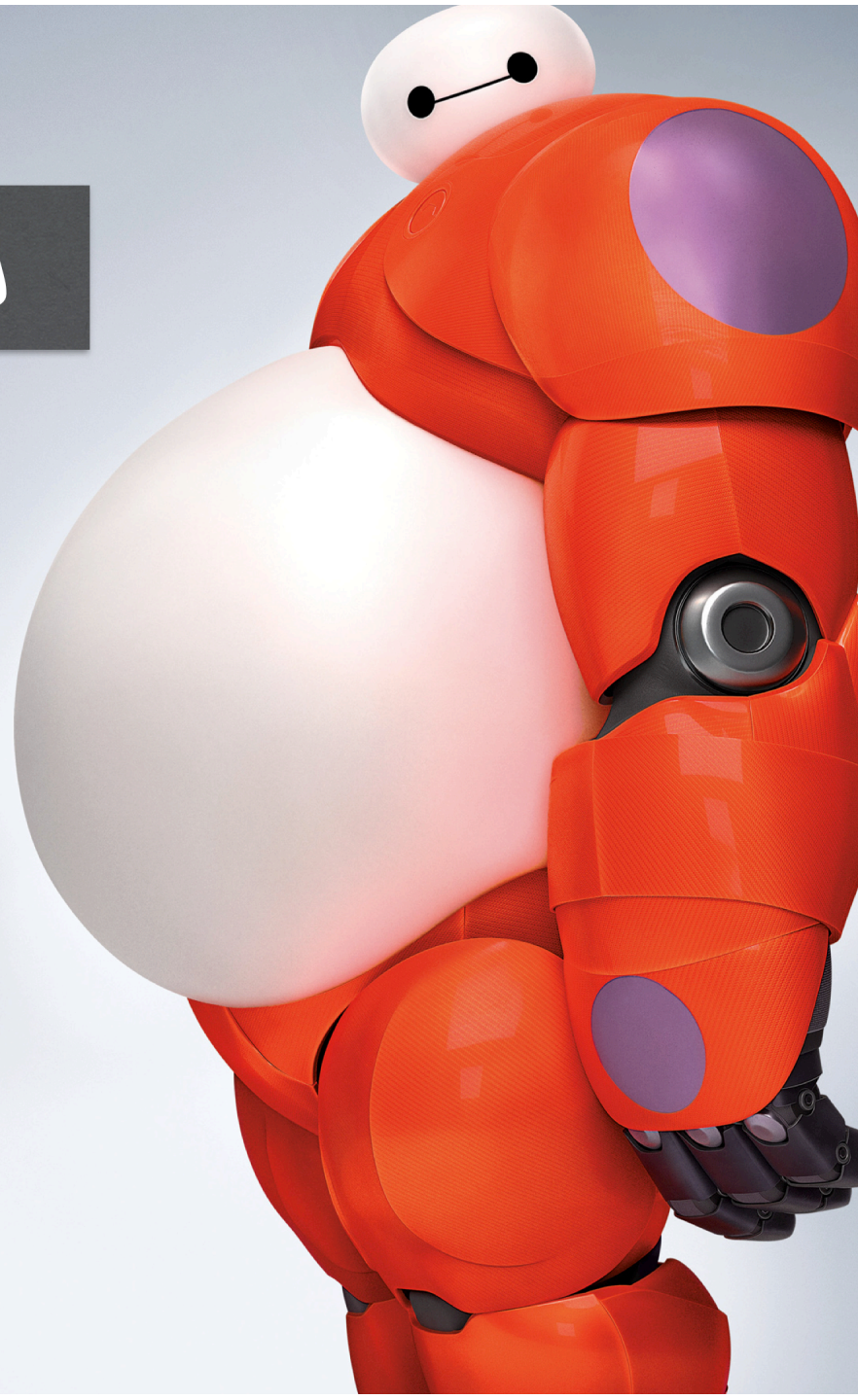
Machine Learning is Everywhere



# Summary



# Human<sub>(bias)</sub> and Algorithms





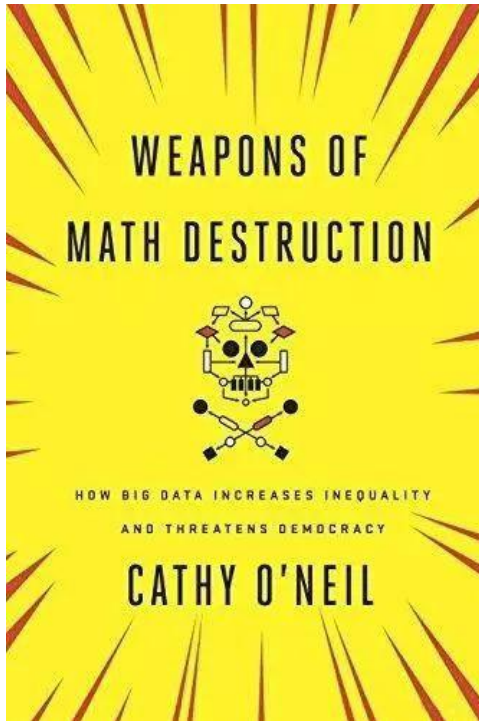
# Two areas of concern: data and algorithms

## Data inputs:

- Poorly selected (e.g., observe only car trips, not bicycle trips)
- Incomplete, incorrect, or outdated
- Selected with bias (e.g., smartphone users)
- Perpetuating and promoting historical biases (e.g., hiring people that "fit the culture")

## Algorithmic processing:

- Poorly designed matching systems
- Personalization and recommendation services that narrow instead of expand user options
- Decision making systems that assume correlation implies causation
- Algorithms that do not compensate for datasets that disproportionately represent populations
- Output models that are hard to understand or explain hinder detection and mitigation of bias



Cathy O'Neil, a mathematician and the author of *Weapons of Math Destruction*, a book that highlights the risk of algorithmic bias in many contexts, says people are often too willing to trust in mathematical models because they believe it will remove human bias.



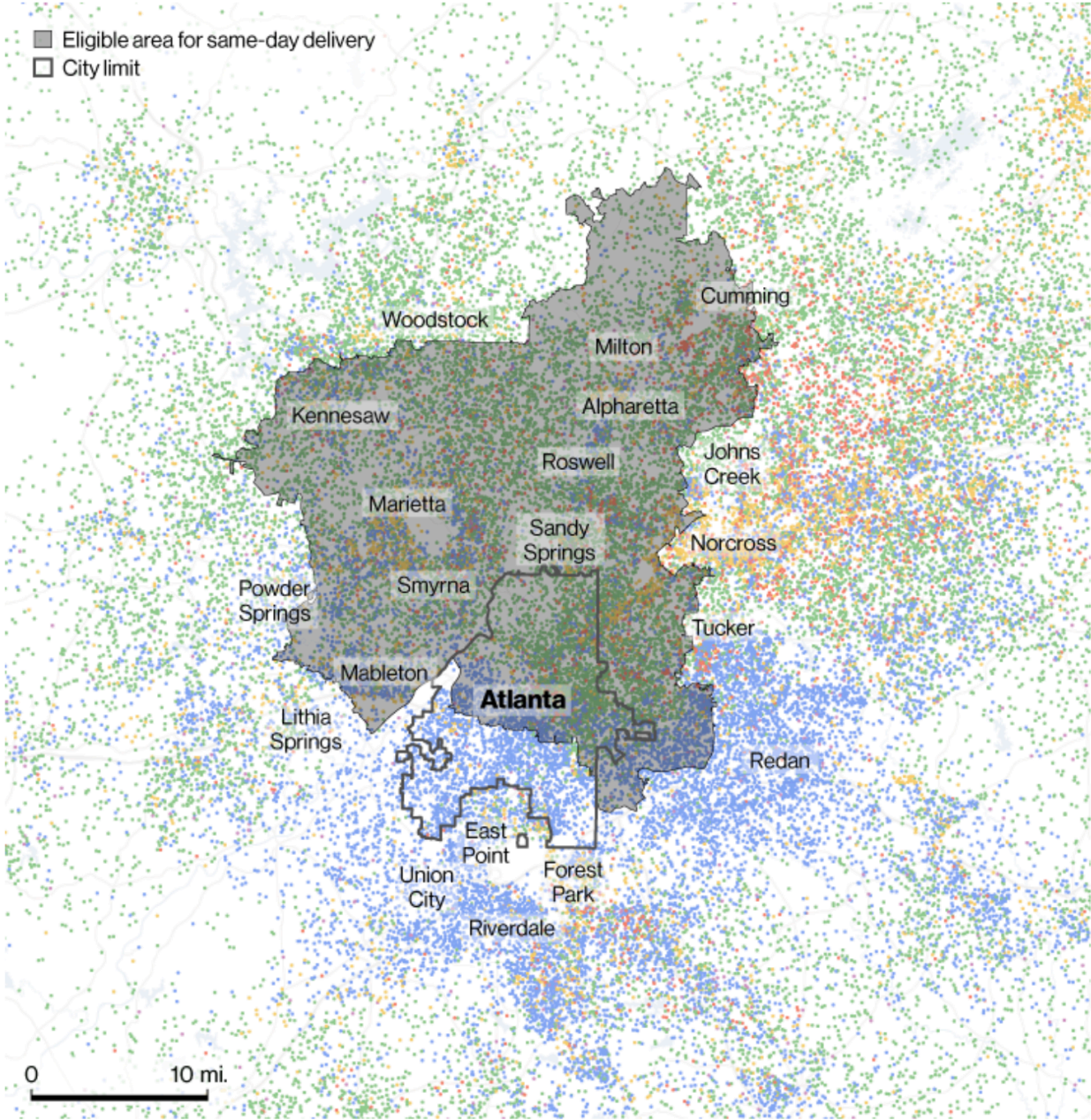
# Judiciary use of COMPAS scores



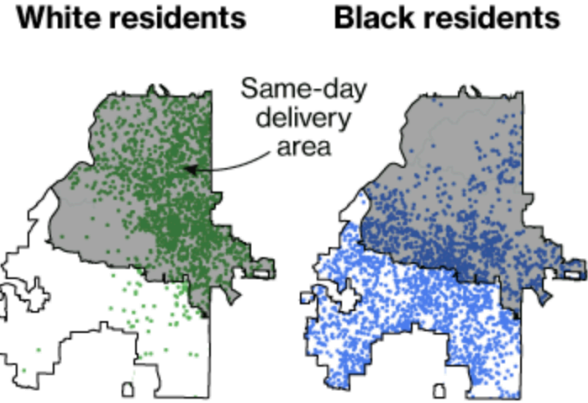
COMPAS (Correctional Offender Management Profiling for Alternative Sanctions):  
137-questions questionnaire and predictive model for "risk of recidivism"

Prediction accuracy of recidivism for blacks and whites is about 60%, but ...

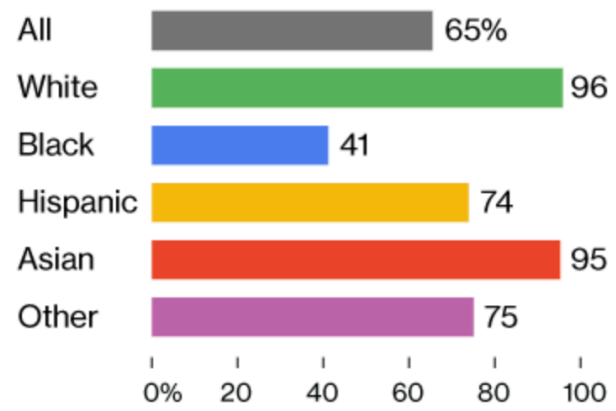
- Blacks that did not reoffend  
were classified as **high risk** twice as much as whites that did not reoffend
- Whites who did reoffend  
were classified as **low risk** twice as much as blacks who did reoffend



The northern half of Atlanta, home to 96% of the city's white residents, has same-day delivery. The southern half, where 90% of the residents are black, is excluded.



**Percentage of residents living in ZIP codes with same-day delivery**



Population percentages are based on American Community Survey estimates and have a 90% confidence interval.



# The ethical challenges

- Algorithmic bias is shaping up to be a major societal issue at a critical moment in the evolution of machine learning and AI.
- If the bias lurking inside the algorithms that make ever-more-important decisions goes unrecognized and unchecked, it could have serious negative consequences, especially for marginalized communities and minorities.



# Some case studies of algorithmic bias

# Unequal Representation and Gender Stereotypes in Image Search Results for Occupations

- Algorithms can be biased in how they represent the world.
- The information people access affects their understanding of the world around them and the decisions they make: biased information can affect both how people treat others and how they evaluate their own choices or opportunities.
- The paper experimentally evaluates the effects of bias in image search results on the images people choose to represent those careers and on people's perceptions of the prevalence of men and women in each occupation.

# Findings

- Stereotype exaggeration: Results for many occupations exhibit a slight exaggeration of gender ratios according to stereotype: e.g., male-dominated professions tend to have even more men in their results
- Systematic over-/under- representation: Search results also exhibit a slight under-representation of women in images, such that an occupation with 50% women would be expected to have about 45% women in the results on average.

# Findings

- Qualitative differential representation: Image search results also exhibit biases in how genders are depicted: those matching the gender stereotype of a profession tend to be portrayed as more professional-looking and less inappropriate-looking.
- Perceptions of occupations in search results: We find that people's existing perceptions of gender ratios in occupations are quite accurate, but that manipulated search results can have a small but significant effect on perceptions, shifting estimations on average ~7%.

# On the web: race and gender stereotypes reinforced

- Results for "CEO" in Google Images: 11% female, US 27% female CEOs
  - Also in Google Images, "doctors" are mostly male, "nurses" are mostly female
- Google search results for professional vs. unprofessional hairstyles for work

Image results:  
"Unprofessional  
hair for work"



Image results:  
"Professional  
hair for work"



# Class Activity 1

# Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment

- Experimental study on Airbnb showing that applications from guests with distinctively African-American names are 16% less likely to be accepted relative to identical guests with distinctively White names.
- Discrimination occurs among landlords of all sizes, including small landlords sharing the property and larger landlords with multiple properties.
- Both African-American and White hosts discriminate against African-American guests; both male and female hosts discriminate; both male and female African-American guests are discriminated against.
- Airbnb's current design choices facilitate discrimination and raise the possibility of erasing some of these civil rights gains.





# Class Activity 2

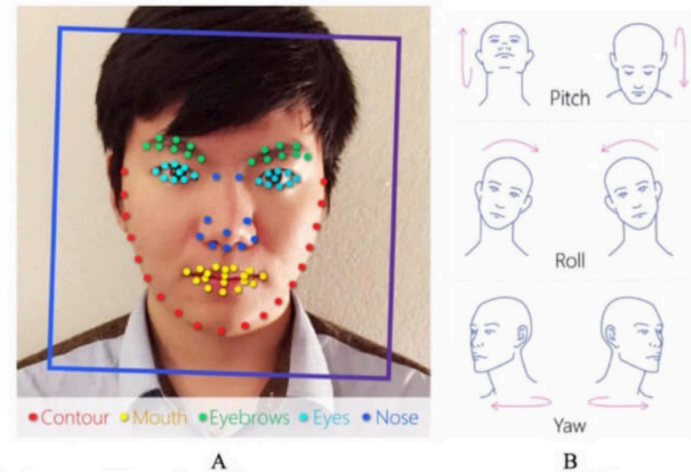
# Deep neural networks are more accurate than humans at detecting sexual orientation from facial images

- Authors used deep neural networks to extract features from 35,326 facial images.
  - Images scraped from public profiles posted on a U.S. dating website
- These features were entered into a logistic regression aimed at classifying sexual orientation.
- Given a single facial image, a classifier could correctly distinguish between gay and heterosexual men in 81% of cases, and in 74% of cases for women.
- The authors claimed that their findings therefore provided “strong support” for the idea that sexual orientation stems from hormone exposure in the womb

# The Study Claiming AI Can Tell If You're Gay or Straight Is Now Under Ethical Review


By Lisa Ryan [@lisarya](#)

SEPTEMBER 12,  
2017  
6:21 PM



An image from the study. Photo: Journal of Personality and Social Psychology/Stanford University

A recent Stanford University study published in the *Journal of Personality and Social Psychology* claimed artificial intelligence can figure out if a person is gay or straight by analyzing pictures of their faces. However, the [Outline](#) reports the study was met with “immediate backlash” from the AI community, academics, and LGBTQ advocates alike — and the paper is now under ethical review.



Some argued that the study is just the latest example of a disturbing technology-fueled revival of physiognomy, the long discredited notion that personality traits can be revealed by measuring the size and shape of a person's eyes, nose and face.



# Class Activity 3

# DeepMind's new AI ethics unit is the company's next big move

Google-owned DeepMind has announced the formation of a major new AI research unit comprised of full-time staff and external advisors



By **JAMES TEMPERTON**

—  
*Wednesday 4 October 2017*

An illustration featuring two stylized, blocky human faces in shades of orange and brown at the bottom. They are looking upwards towards a dense field of colorful gears in red, yellow, and teal against a black background. Dashed lines and small white crosses connect the gears, suggesting a complex network or system.

Google DeepMind



# Job Openings

## Artificial Intelligence/FutureTech Investigative Reporter

📍 New York, NY

Apply

Apply with LinkedIn

🕒 Posted 30+ Days Ago

📄 Full time

📄 REQ-001480

### Job Description

Investigate how algorithms, artificial intelligence, robots and technology are influencing our lives, our businesses, our privacy and the future.

This deeply-informed reporter will be able to understand and explain complex technologies while investigating the people and companies behind them. They will be expected to discover and cultivate sources and contacts and to break ground reporting on issues that many companies would rather go uncovered. They will also be comfortable with - and even capable of - a variety of computer-assisted reporting techniques. The reporter will work on a small team and be interested in telling stories through multiple mediums including interactive graphics, virtual reality, audio, video and of course the written word.

### About Us



### Help shape the future Times

This is an important moment to v organization, we're taking advant landscape to pioneer a new era o original reporting at our core, we' about our reader relationships an vant offerings and experiences. V

Location is flexible

# New York Times

*This is a guild position.*

To apply:

# Research Attention

- In 2017, a group of researchers, together with the American Civil Liberties Union, launched an effort to identify and highlight algorithmic bias, called AI Now



F.A.T



Fairness



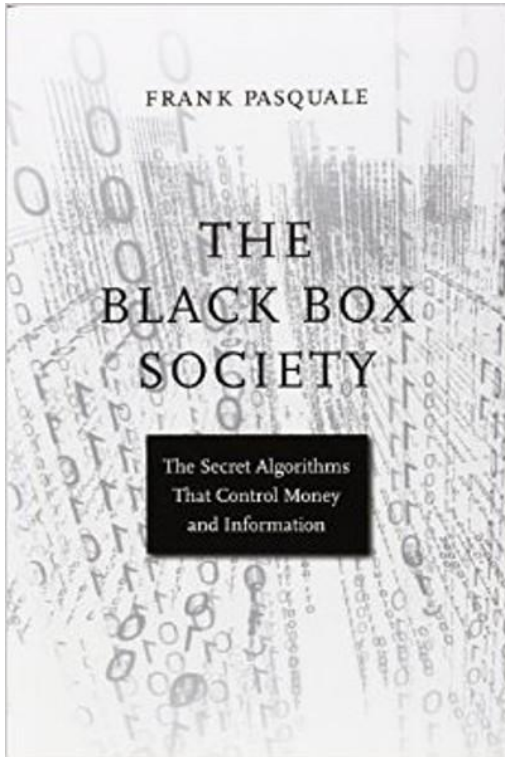


Accountability





Transparency



Algorithms are "black boxes" protected by

Industrial secrecy

Legal protections

Intentional obfuscation

Discrimination becomes invisible

Mitigation becomes impossible

*F. Pasquale (2015): The Black Box Society. Harvard University Press.*



MAY  
2018

HISTORY

Right to be Forgotten



MAY  
2018



Right to Explanation