UCSD - COGS 108 - DATA SCIENCE IN PRACTICE

DATA SCIENCE ETHICS

THOMAS DONOGHUE - FEBRUARY 1ST, 2018

- Newsfeed Propaganda [link]
- Racist Soap Dispenser [video]
- Delivery regions [link]
- Face-identification / etc. [link]
- Video platform labelling LGBT content [link]
- Online classes [link]
- Translation app [link, link, paper]

ON POWER & OBJECTIVITY

- Data Science is powerful.
- Data, algorithms and analysis are not objective.

ETHICAL DATA SCIENCE

- Data science that is pursued in a manner so that is equitable, with respect for privacy and consent, so as to ensure that it does not cause undue harm.

ETHICALLY BETTER DATA SCIENCE IS (OFTEN) TECHNICALLY BETTER DATA SCIENCE

NOT ALL QUESTIONS ARE DATA SCIENCE QUESTIONS

DEFINITIONS

- Normative Ethics: What One *Ought* To Do
- Utilitarian: Actions are right if they benefit the majority
- Rules / Rights Based: Actions are right so far as they don't violate rules and/or others fundamental rights
- Blameworthiness: whether one is blameworthy for wrongdoing (regardless of intent)

ON THE LAW

- There are laws that cover research, privacy, and other aspects of data use / data science.
- Regulations are needed.
- However: this talk is not really about that.

8 FUN WAYS TO NOT RUIN PEOPLES LIVES WITH DATA SCIENCE

1] THE QUESTION

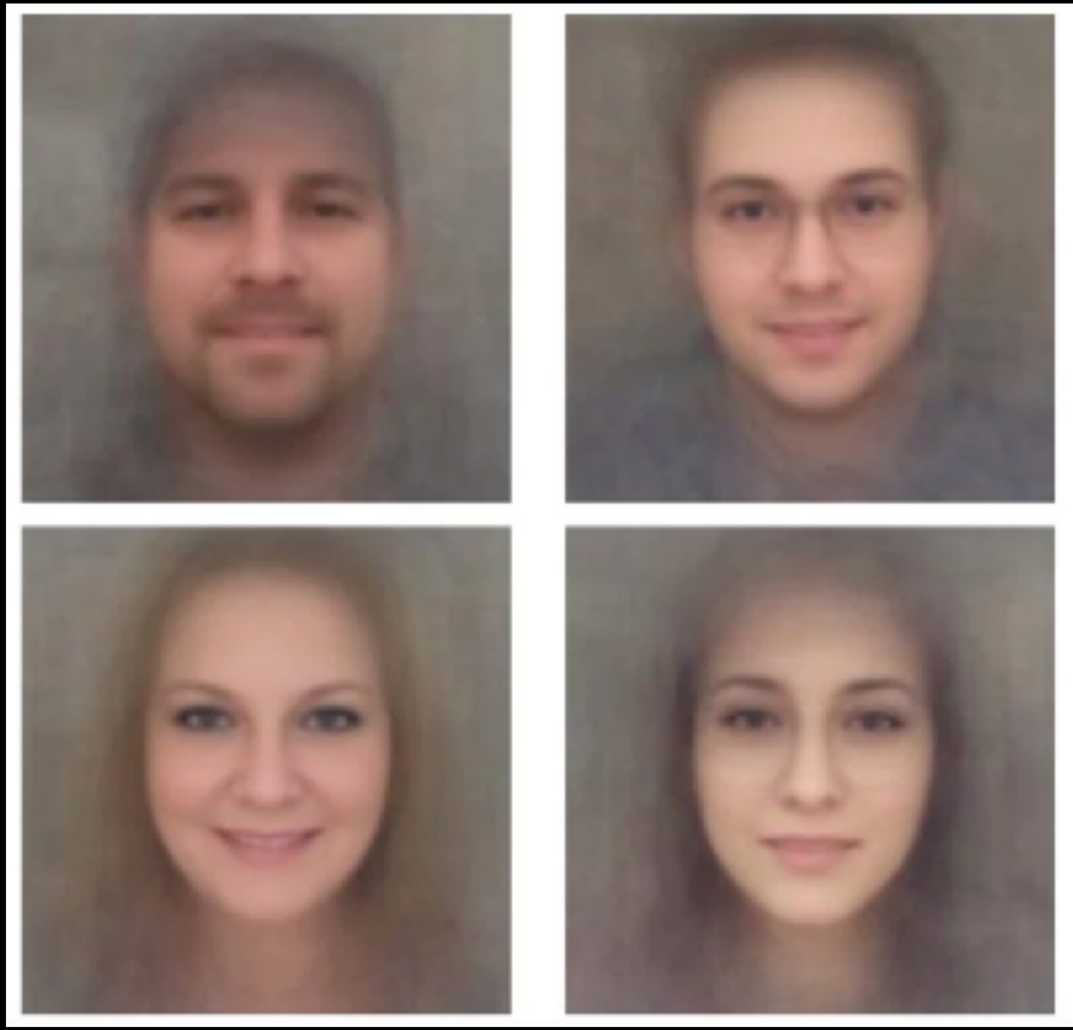
- What is your question? Is it well posed?
- Do you know something about the context and background of your question?
- What is the scope your investigation? What correlates might you inadvertently track?

CASE STUDY: LABELLING FACES

- Detecting Criminality from Faces
- Source: link, paper
- Detecting Sexual Orientation from Faces with Computer Vision
- Source: link, paper

(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n



2] THE DATA

- Is there data available? Is this data directly related to your question, or only potentially related through proxies?
- Who do you have data from?
- Do you have enough data to make reliable inferences?
- What biases does your data have?
- If you do not have, and can not get, enough good, appropriate data, you may just have to stop.

CASE STUDY: BIOMEDICAL SCIENCE

- Biomedical research has often excluded female subjects

- This was based on a (faulty) assumption that females would be more variable
- These findings do not generalize as well
- Sources: link, link, link.

ASIDE: RESEARCH ETHICS

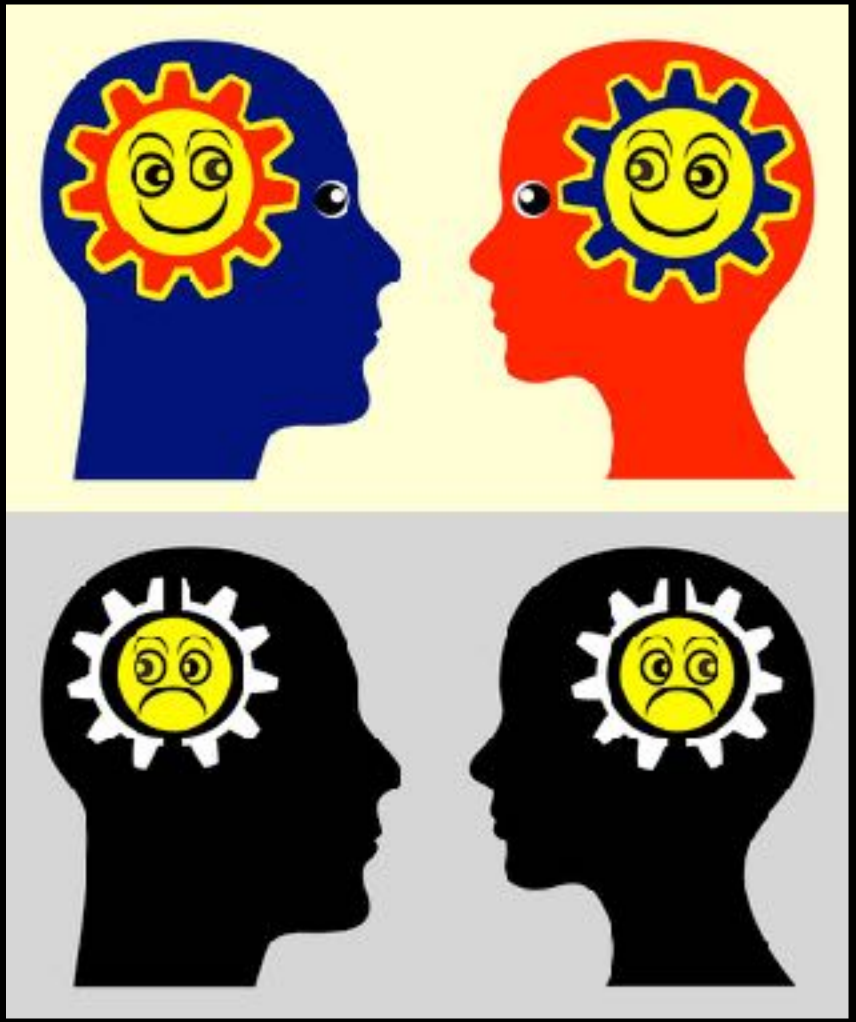
- Research: a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge.

3] INFORMED CONSENT

- Informed Consent: the voluntary agreement to participate in research, in which the subject has an understanding of the research and its risks.

CASE STUDY: EMOTIONAL CONTAGION

- Facebook conducted an experiment investigating whether they could manipulate peoples emotions by selecting the content of the newsfeed.
- Source: link, paper



4] PRIVACY

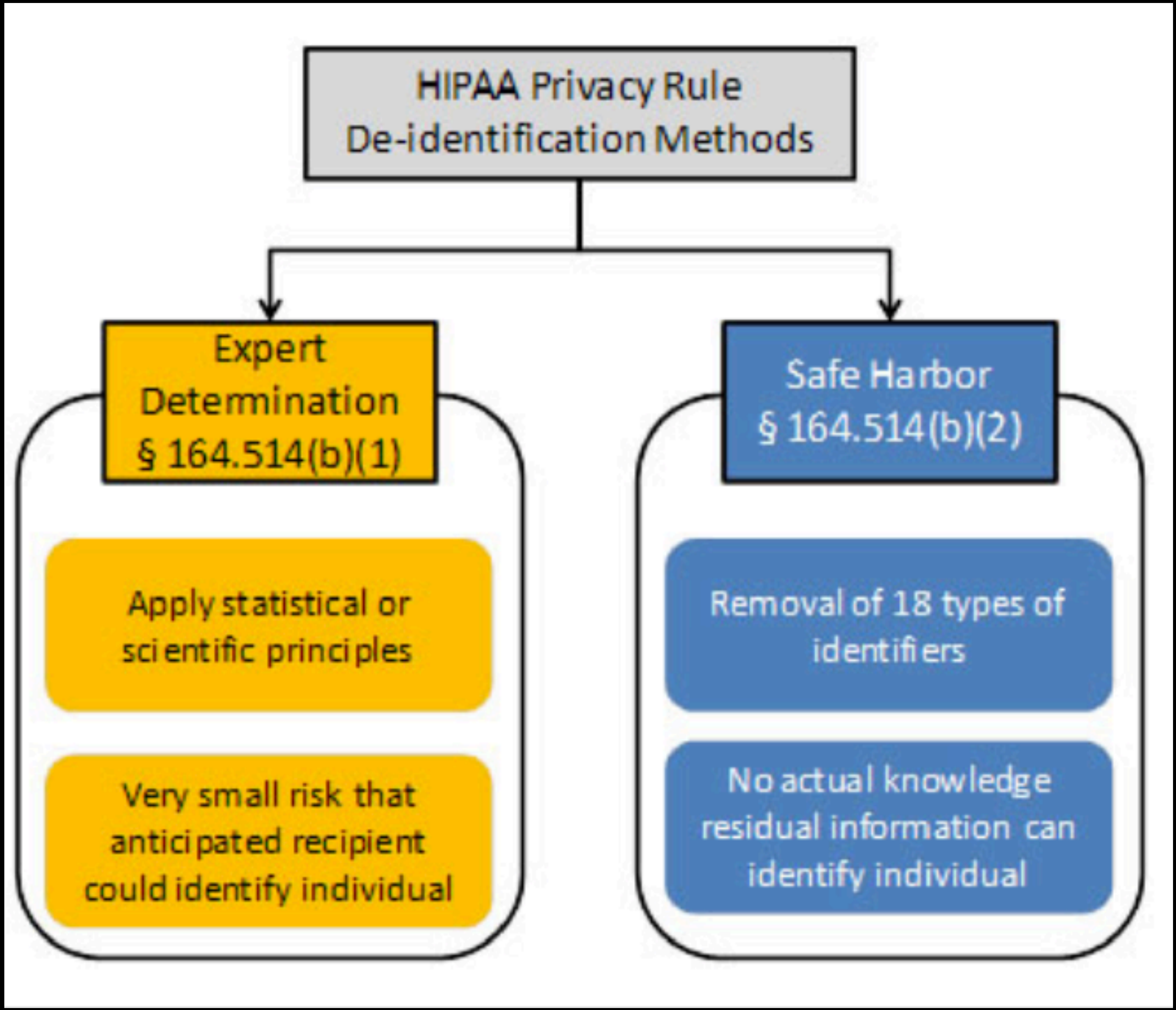
- Can you guarantee privacy?
- What is the level of risk of your data, and how will you mitigate the risks? Are all subjects equally vulnerable?
- Anonymization: the process of removing personally identifiable information from data sets.
- Use secure data storage, with appropriate access rights

CASE STUDY: RUNNING DATA

- The running company Strava released running data, geotagged from around the world.
- Routes are clear around sensitive locations, including military bases.

SAFE HARBOUR METHOD

- A method that specifies personally identifiable data to remove from a dataset for the purpose of de-identification.

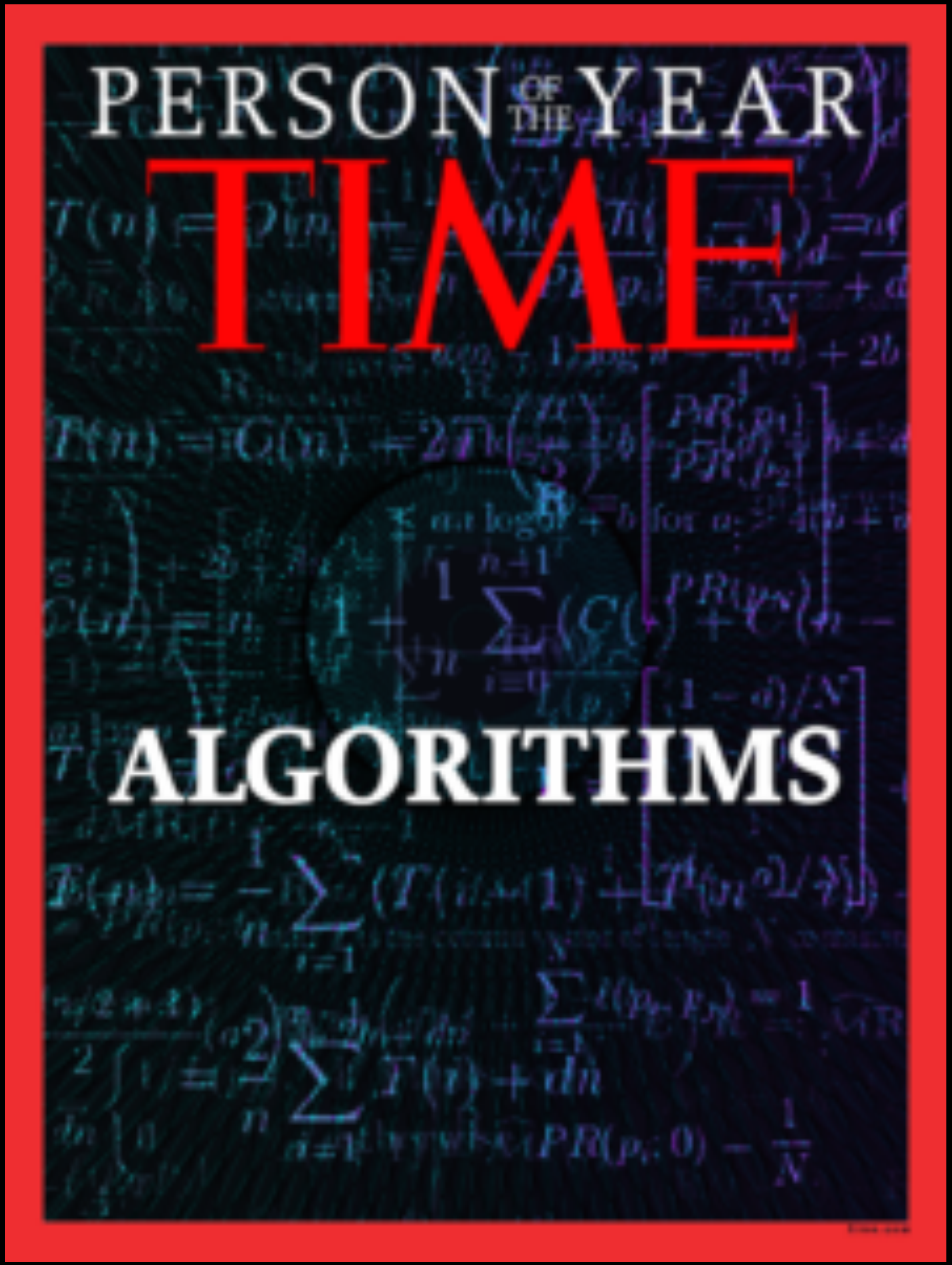


5] EVALUATION

- How will you evaluate the project?
  - Do you have a verifiable metric of success?
- Is it a self-fulfilling prophecy?

CASE STUDY: TEACHER RATING

- Washington DC school district used an algorithm to rate teachers, based on test scores. Scores from this algorithm were used to fire 'low-performers'
- They had no independent measure of whether this measure improved teaching.



6] ANALYSIS

- Do your analyses reflect spurious correlations?
  - Can you tease apart causation?
- What kind of covariates might you be tracking?
  - Are you inferring latent variables from proxies?

7] TRANSPARENCY & APPEAL

- Is your model a black box?
  - Is it interpretable as to how it came to any particular decision?
- Is there a way to appeal a model decision?
  - What kind of evidence would you need to refute a decision?

CASE STUDY: PREDICTIVE POLICING

- Predictive policing uses algorithms to predict crime, and recidivism
  - Input data can be highly correlated with race & SES, reflecting spurious correlations and leading to discriminatory decisions.
- These algorithms and decisions are often opaque, and un-appealable.

8] CONTINUOUS MONITORING

- Healthy models maintain a back and forth with the the thing(s) in the world they are trying to understand.
- Are you tracking for changes related to your data, assumptions, and evaluation metrics?
- Are you proactively looking for potential un-intended side effects, or harmful outputs?

CASE STUDY: NEWS SHARING

- Facebook is continuously making predictions about what you are going to do, which it uses to try to influence behaviour, and then update it's models based on the results.

- These models optimize for engagement & sharing in such a way that can be gamed, and promote the spreading of misinformation.

ON SYSTEMS & INCENTIVE STRUCTURES

- Novel systems are not, de facto, equalizers. They will tend towards propagating existing inequalities.

- Companies working on these systems may have conflicts of interest with respect to the incentive structures imposted by the system and/or the business

ON PERPETUATING INEQUALITY

- Data & Algorithms can & will entrench social disparities
- Errors and bias typically target the disenfranchised
- The combination of damage, scale, and opacity can be incredibly destructive
- They can introduce feedback in such a way as to enact self-fulfilling prophecies

PULLING IT ALL TOGETHER (GOOD)

- You have a well-posed & scoped question, that you know something about.
- There are adequate, data covering the population of interest, with known and manageable biases.
- You are allowed to use this data for these purposes.
- You use appropriately de-identified data, stored securely.
- You have defined metrics for success, objectively measured.
- You push your analysis to establish causality.
- Your model is understandable, and/or you have a procedure to appeal.
- You monitor the system for changes related to your data, underlying assumptions, outcomes, & impact.

HOW TO BE (ACCIDENTALLY) BAD

- You have an ill-posed question on a topic you are totally unfamiliar with.
- You use haphazardly collected, biased data, from a subset of the population.
- You did not check whether you are allowed to use the data for this project.
- You use un-anonymized personally identifiable data, stored insecurely.
- You have no clear metric for success, other than if 'it seems to work'.
- You un-critically make use of spurious correlations (especially for proxies).
- Your model is a black box, with no tractable way to appeal.
- You do not monitor for potential changes related to your data, model performance, underlying assumptions, or outcomes & impact.

CASE STUDIES

- Automated labelling based on faces
- Exclusion of female subjects in biomedical research
- Emotional Contagion
- Privacy of Running Data
- Algorithmic Teacher Evaluation & Dismissal
- Predictive Policing
- News sharing

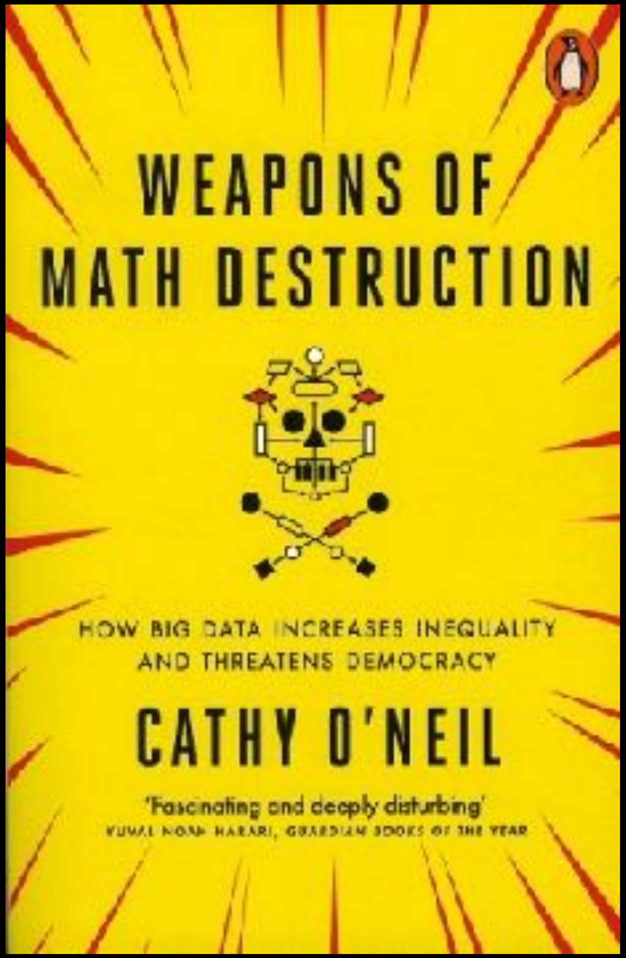
- Newsfeed Propaganda [link]
- Racist Soap Dispenser [video]
- Delivery regions [link]
- Face-identification / etc. [link]
- Video platform labelling LGBT content [link]
- Online classes [link]
- Translation app [link, link, paper]

MISCELLANEOUS POINTS

- Data Science is a tool.
- Data Science is an interdisciplinary, team sport.

ACKNOWLEDGEMENTS & RESOURCES

- Thank you to those who helped me create this talk
- Join the conversation about data science ethics



DATA FOR DEMOCRACY