Data Science Ethics

Jason G. Fleischer, Ph.D
UC San Diego

Department of Cognitive Science
jfleischer@ucsd.edu
https://jgfleischer.com
@jasongfleischer

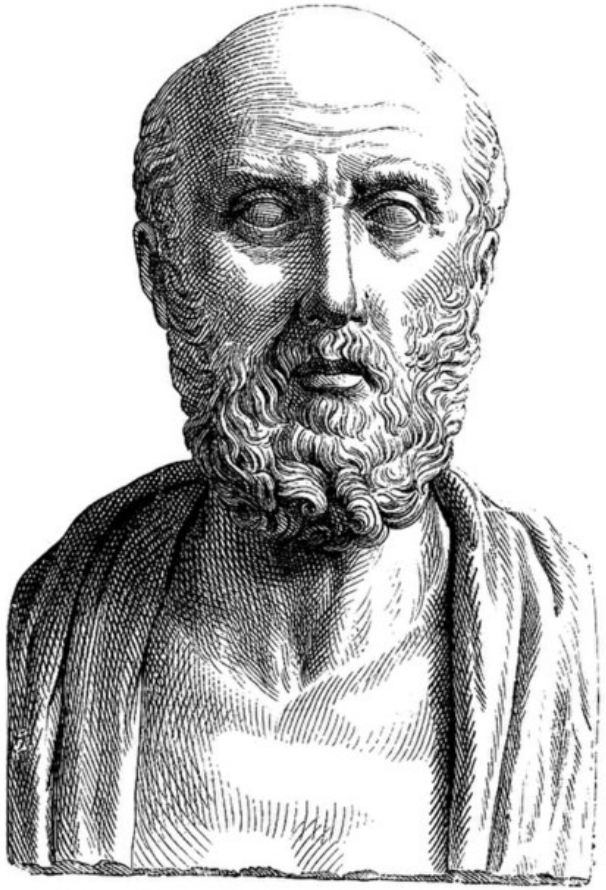
Course Reminders

- Due Today
  - FinAid survey on Canvas
  - D1
  - Last chance to fill in the pre-course survey!!
- Due Monday
  - Q2
  - Github ID
  - Group signup
- Due Wednesday
  - A1

ETHICS

“Moral principles that govern a person's behaviour or the conducting of an activity.”

Source: Oxford Dictionary

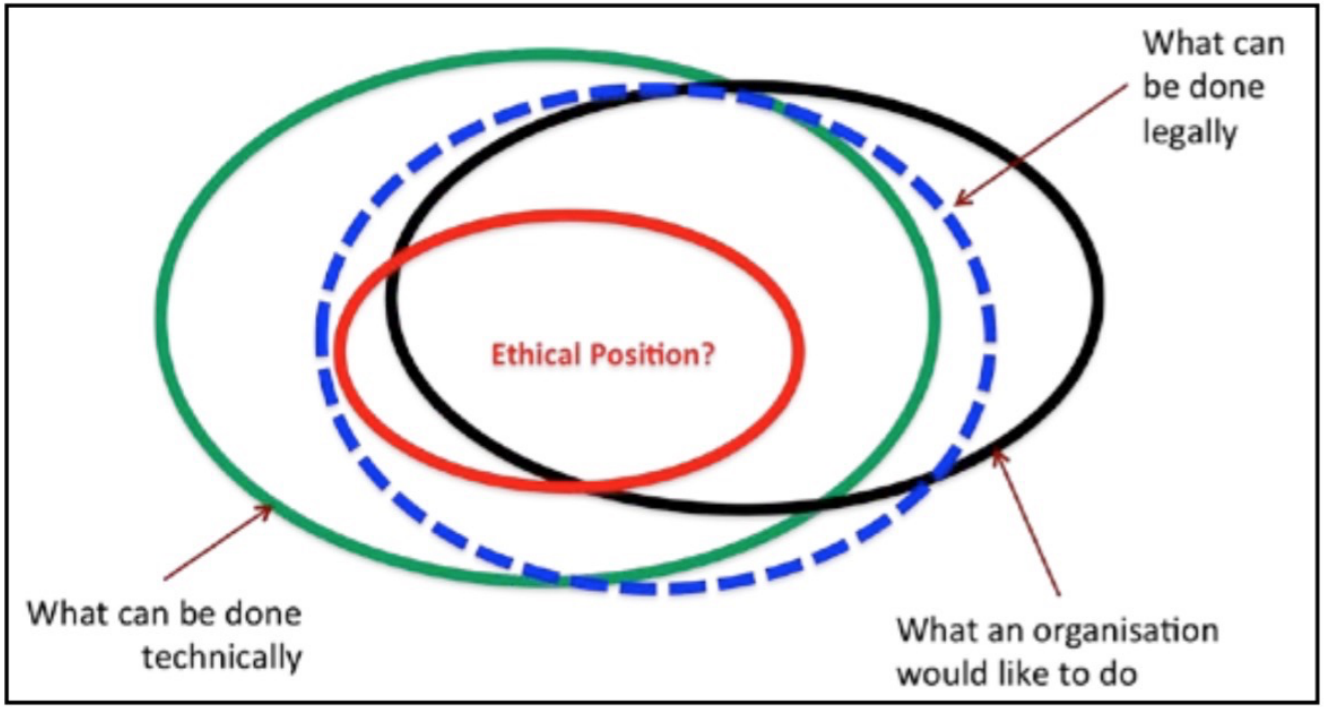


Professional ethics

The Hippocratic Oath (Medicine)

- First do no harm
- Preserve patient privacy
- Do not be afraid to say I don’t know
- Obligations to ALL human beings

Data Ethics



Source: IBM

adapted from Brad Voytek

Data Science Ethics

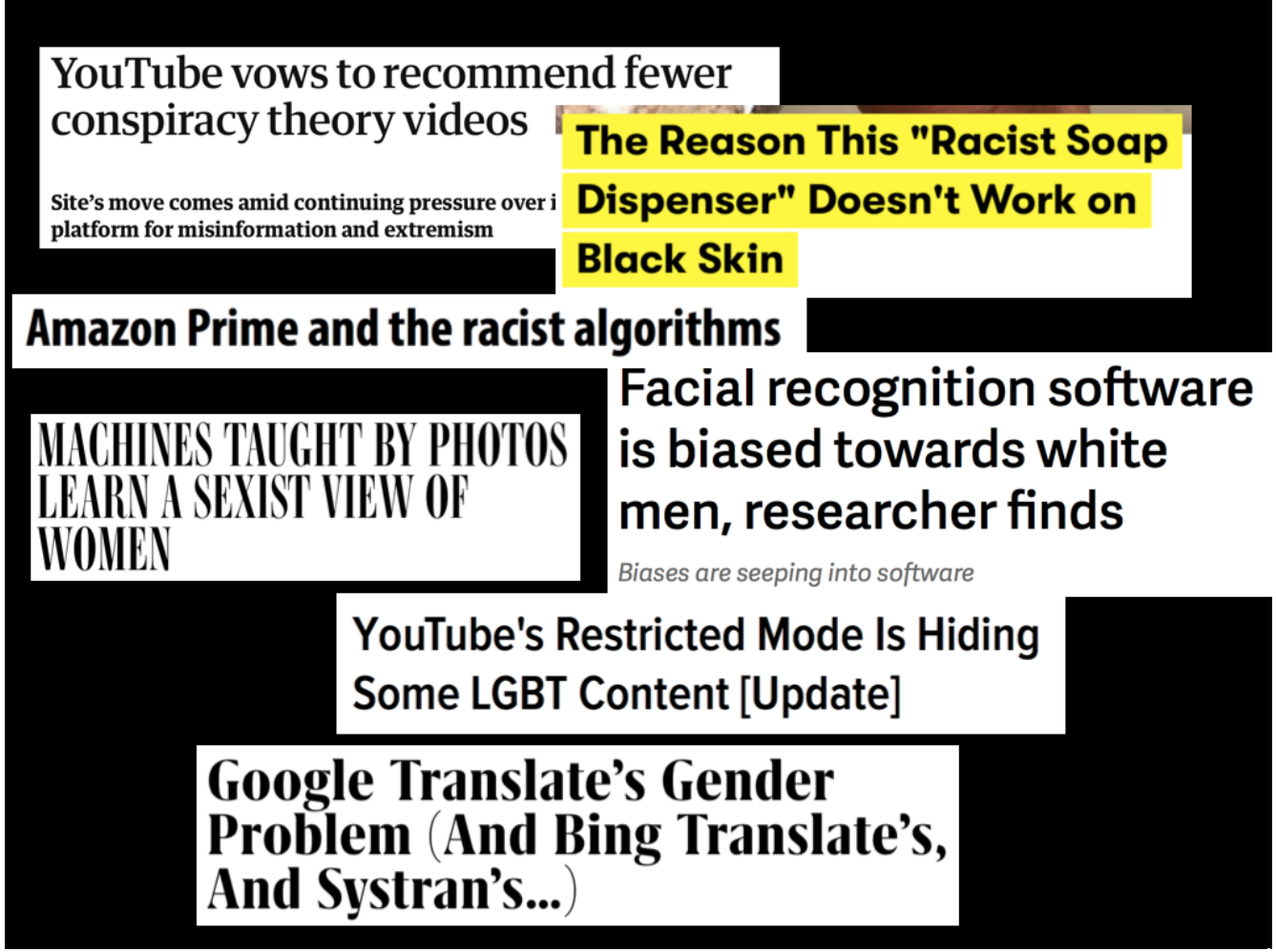
https://forms.gle/iysz1P1JMyJWiL4p6

Ethical Data Science

Data science 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.

On INTENT and OBJECTIVITY

- Intent is not required for harmful practices to occur
- Data, algorithms and analysis are not objective.
  - It is done by people, who have biases
  - It uses data, which have biases
- Data Science is powerful
- Bias & discrimination driven by data & algorithms can give new scale to pre-existing inequities



adapted from Thomas Donoghue

COGS 9 Examples

- Ashley Madison Hack [link]
- OKCupid Data Published [link]
- Equifax Hack [link]
- Google & Pentagon Team Up on Drones [link]
- Cambridge Analytica Data Breach To Influence US Elections [link]
- Amazon and Police Team Up on Facial Recognition & Surveillance [link]
- Amazon scraps secret AI recruiting tool biased against women [link]

A few additional examples I've compiled in the last 2 years…

- Study of bias in AI [link]
- Pasco County Algorithmic Bias [link]
- Ethical issues (misogyny, racism) in large available datasets [link],[link]
- Florida COVID-19 dashboard data scientist debacle [link]
- Banjo surveillance via fake apps [link]
- Google fires AI ethics founder [link] & Timnit Gebru's firing [link]
- Twitter fires entire ethics & compliance team [link]
- MS lay off their entire ethics teams [link]
- ChatGPT is dumber than you think [link]
- Synthetic Media Creates New Social Engineering Threats [link]
- Generative AI art is a copyright nightmare [link]

NINE THINGS TO CONSIDER TO NOT RUIN PEOPLE'S LIVES WITH DATA SCIENCE

1. THE QUESTION
2. THE IMPLICATIONS
3. THE DATA
4. INFORMED CONSENT
5. PRIVACY
6. EVALUATION
7. ANALYSIS
8. TRANSPARENCY & APPEAL
9. CONTINUOUS MONITORING

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? Is it possible to answer this question well?

Media file

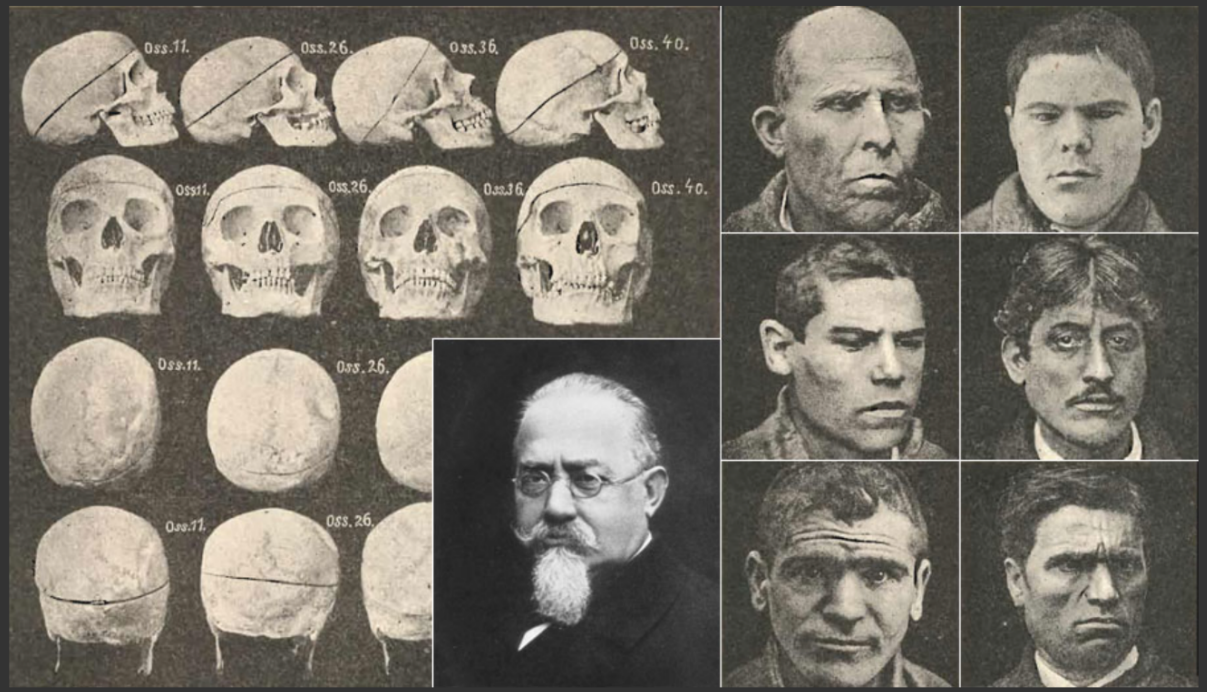
Racial Photograph

View media page | View source file

Citations of this media

"Origin of Criminology
From its inception photography had a profound effect on anthropological work as measuring device when studying races. While many would expect the photograph to bring truth and an end to the accentuated stereotyping by hand-drawn image, the medium would still be used to promote ideas of racial inferior traits. For example, Carl Victor and Friedrich Wilhelm Dammann's photographic book, Races of Men has influenced and propelled the viewpoints and stereotypes of different races. Containing black and white photos along with brief captions describing physical and mental traits, the context of these depictions serve to relay the idea of a Darwinian racial evolution from the Polynesians culminating with the Germanic race. Alphonse Bertillon founded modern anthropometric photography for the purpose of identifying repeated offenders by photographing and recording measures of physical features that remain constant throughout an individual's adult life. Cesare Lombroso, the founder of anthropological criminology, claimed to identify a links between common physical and mental traits and those highly likely to commit crimes. Dubbing the concept of being a "born criminal" Lombroso argued in favor of biological determinism, He found that skull and facial features were clues to genetic criminality and could be measured into quantitative research. The image depicts some of the 14 traits of a criminal Lombroso identified as large jaws, forward projection of jaw, low sloping forehead; high cheekbones, flattened or upturned nose; handle-shaped ears; hawk-like noses or fleshy lips; hard shifty eyes; scanty beard or baldness; insensitivity to pain; long arms, and so on. Lombroso viewed criminality as a hereditary disposition due to having traits similar to primitive human ancestors of monkeys and apes. His theories have also helped with influencing eugenics and anti-miscegenation laws, while his legacy can be found in modern day policing with racial profiling. "

—from "The Origins of Criminology"



Details

Scalar URL	https://scalar.usc.edu/works/measuring-prejudice/media/racial-photograph (version 1)
Source URL	https://scalar.usc.edu/works/measuring-prejudice/media/racial%20photography.jpg (image/JPEG)
dcterms:title	Racial Photograph
View as	RDF-XML, RDF-JSON, or HTML

Case Study: Labelling Faces

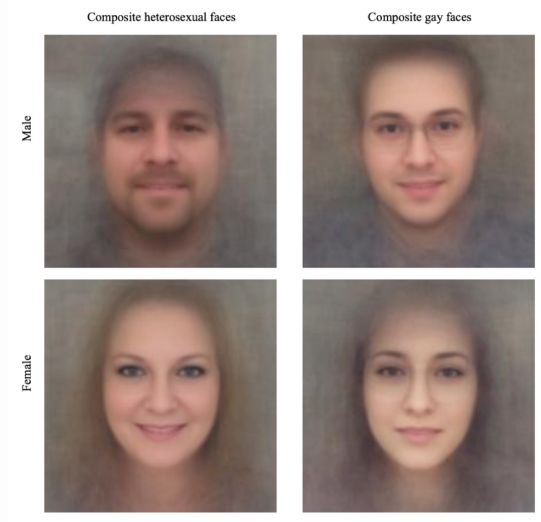
Detecting criminality from faces [link, paper]

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

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

Figure 1. Sample ID photos in our data set.

Detecting Sexual Orientation From Faces with computer vision [link, paper]



This stuff just doesn't go away...

ARTICLE

https://doi.org/10.1038/s41467-020-18566-7 OPEN

Tracking historical changes in trustworthiness using machine learning analyses of facial cues in paintings

Lou Safra [1,2,3]✉, Coralie Chevallier[1], Julie Grèzes[1] & Nicolas Baumard [2]✉

Received: 19 May 2019; Accepted: 10 August 2020;
Published online: 22 September 2020

https://www.nature.com/articles/s41467-020-18566-7.epdf

2. THE IMPLICATIONS

- Who are the stakeholders? How does this affect them?
- Could the information you will gain and/or the tool you are building be co-opted for nefarious purposes?
  a. If so, can you protect them from that?
- Have you considered potential unintended consequences?

Case Study: Abuse of social networks

Facebook has been co-opted by military personnel to spread misinformation, hate speech, and promote ethnic cleansing [news link, UN Report]

3. 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

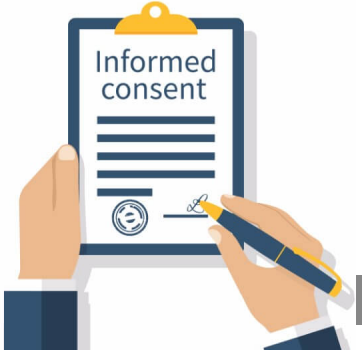
These findings do not generalize as well

Sources: link, link, link

4. INFORMED CONSENT

INFORMED CONSENT: the voluntary agreement to participate in research, in which the subject has an understanding of the research and its risks

Informed consent can be withdrawn at any point in time



Case Study: Biomedical Science

Medical doctors have a history of playing God. Egregiously unethical medical research was famously conducted by Nazis, but also by Americans (Tuskegee Syphilis Study, Chester Southam injecting people with cancer, and many others) and other nations throughout history. This led to the creation of the Belmont report and our current system of IRBs (institutional ethics review boards) for research that involves human subjects.

The Belmont report establishes principles that must be fulfilled for research on humans:

Respect for persons. This principle includes both respect for the autonomy of human subjects and the importance of protecting vulnerable individuals.

Beneficence. More than just promotion of well-being, the duty of beneficence requires that research maximize the benefit-to-harm ratio for individual subjects and for the research program as a whole.

Justice. Justice in research focuses on the duty to assign the burden and benefits of research fairly.

Sources: link, link, link

5. 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 datasets (PII)
- Use secure data storage, with appropriate access rights

Case Study: Running Data

Strava, a company who made an app that released running data, geotagged from around the world [link]

Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Latest: Strava suggests military users 'opt out' of heatmap as row deepens

A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap

Consumer Tech

Don’t sell my data! We finally have a law for that

You’re going to have to jump through some hoops, but you can ask companies to access, delete and stop selling your data using the new California Consumer Privacy Act - even if you don’t live in California.

By **Geoffrey A. Fowler**

FEBRUARY 19, 2020

Our version of Europe’s GDPR law

https://www.washingtonpost.com/technology/2020/02/06/ccpa-

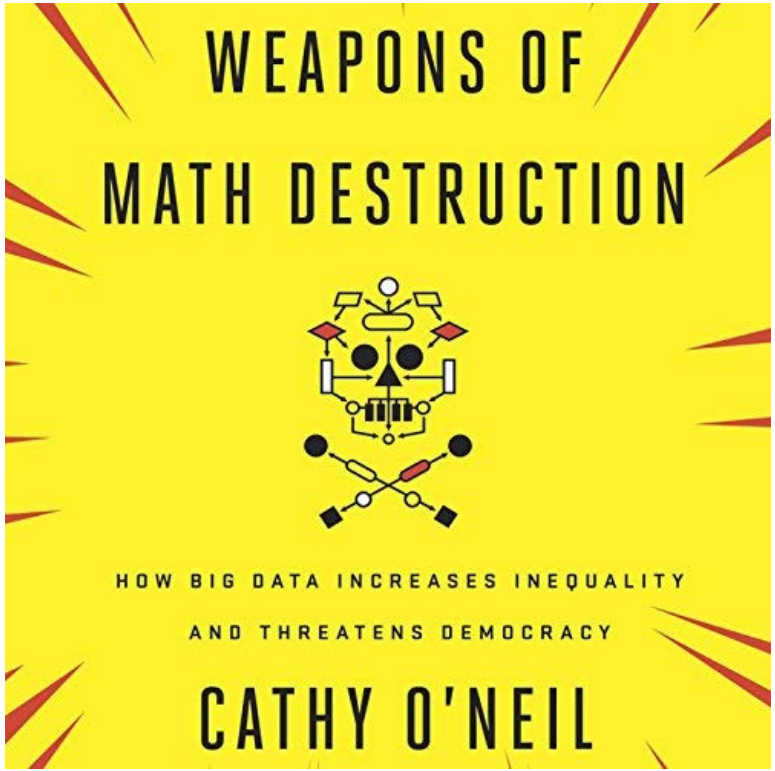
6. EVALUATION

- How will you evaluate the project?
  a. Do you have a verifiable metric of success?
- Goodhart's Law: when a measure becomes a target, it ceases to be a good measure.

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



7. ANALYSIS

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

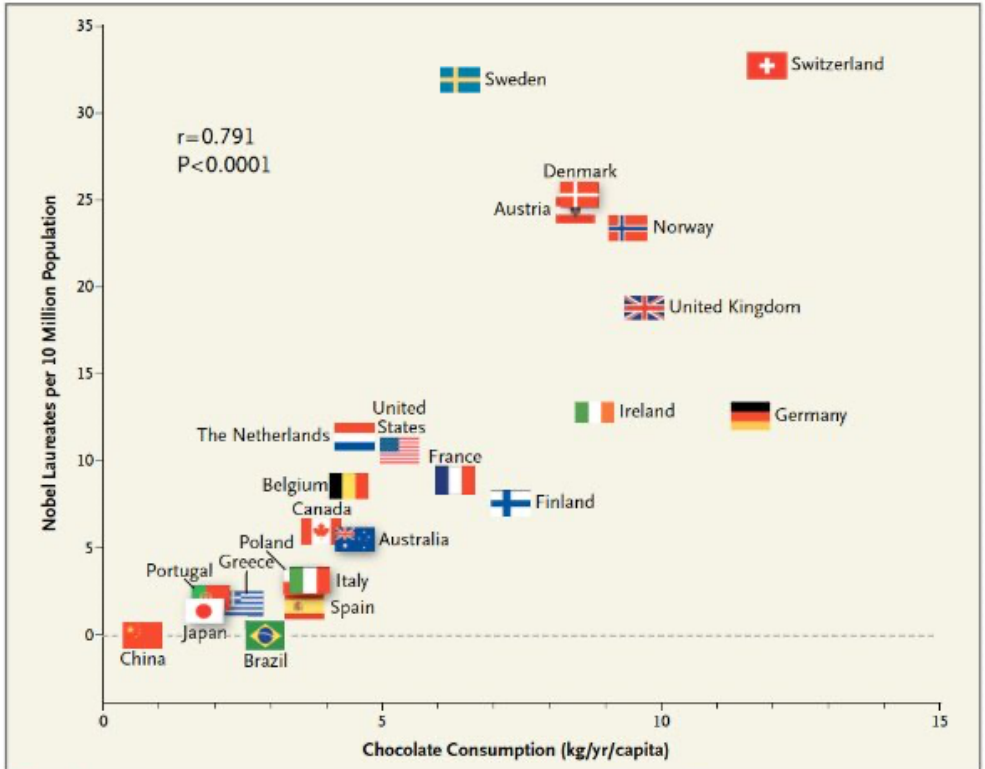
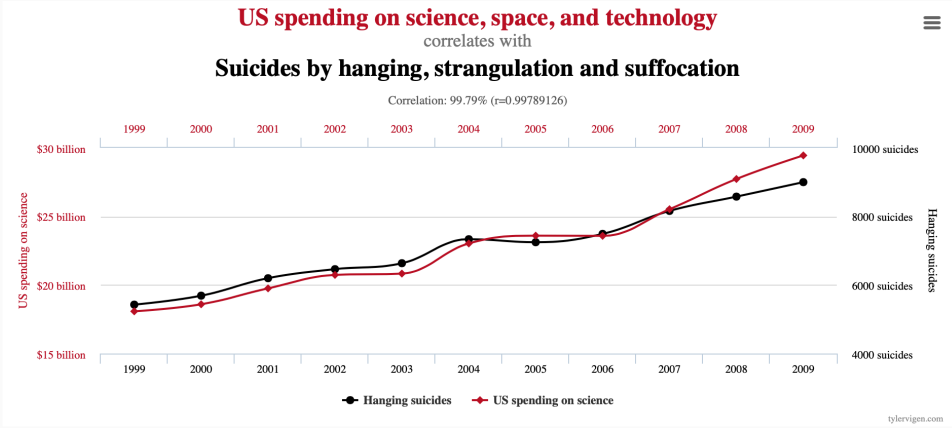


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

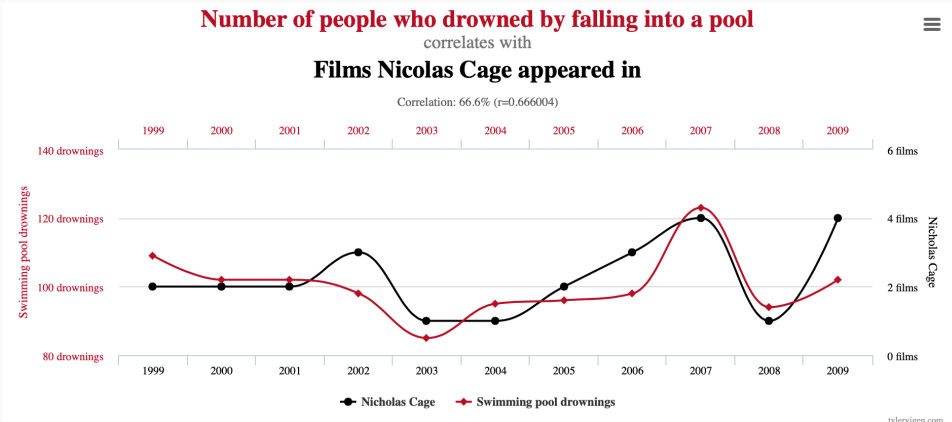
https://www.nejm.org/doi/full/10.1056/NEJMon1211064

Spurious correlations

https://www.tylervigen.com/spurious-correlations



Data sources: U.S. Office of Management and Budget and Centers for Disease Control & Prevention

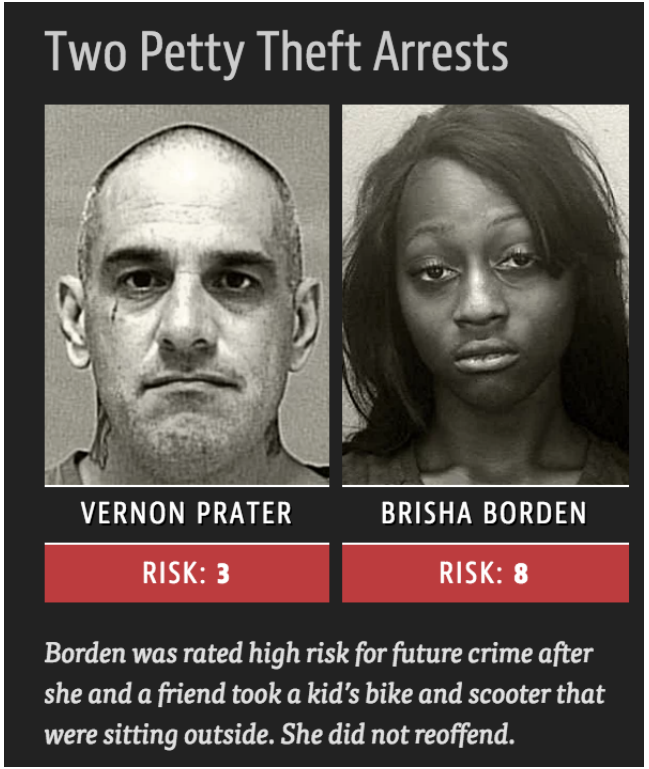


8. TRANSPARENCY & APPEAL

- Is your model a black box?
  a. Is it interpretable as to how it came to any particular decision?
- Is there a way to appeal a model decision?
  a. 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 [link] with race & SES, reflecting spurious correlations and leading to discriminatory decisions.
- These algorithms and decisions are often opaque and un-appealable.



9. CONTINUOUS MONITORING

- Healthy models maintain a back and forth with 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 unintended side effects of your model itself or harmful outputs?
- Do you have a mechanism to fix and update your algorithm?

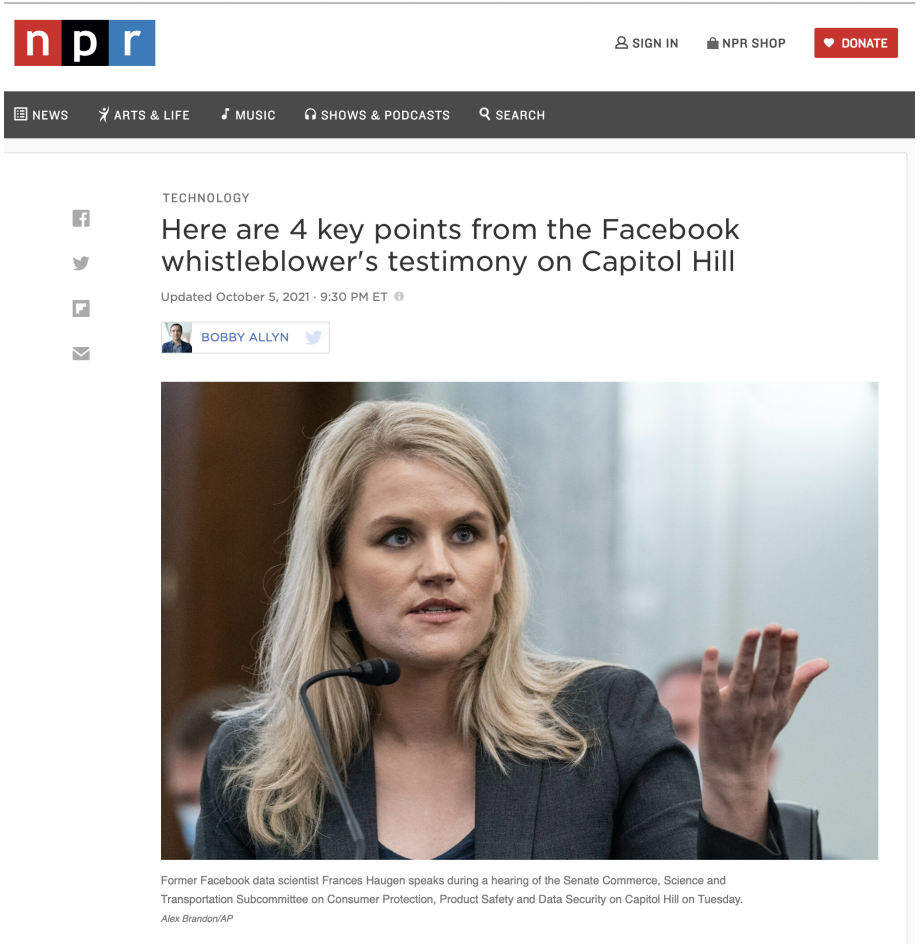
Case Study: Fake news and video reccs

- Companies are continuously making predictions about what you are going to do, which it uses to try to influence behaviour and then update its models based on the results
- Models optimize for engagement and sharing - can promote the spreading of misinformation

13.5% of U.K. teen girls in one survey say their suicidal thoughts became more frequent after starting on Instagram.

Another leaked study found 17% of teen girls say their eating disorders got worse after using Instagram.

About 32% of teen girls said that when they felt bad about their bodies, Instagram made them feel worse



https://www.npr.org/2021/10/05/1043377310/facebook-whistleblower-frances-haugen-congress

Prompt: ceo;
Date: April 6, 2022

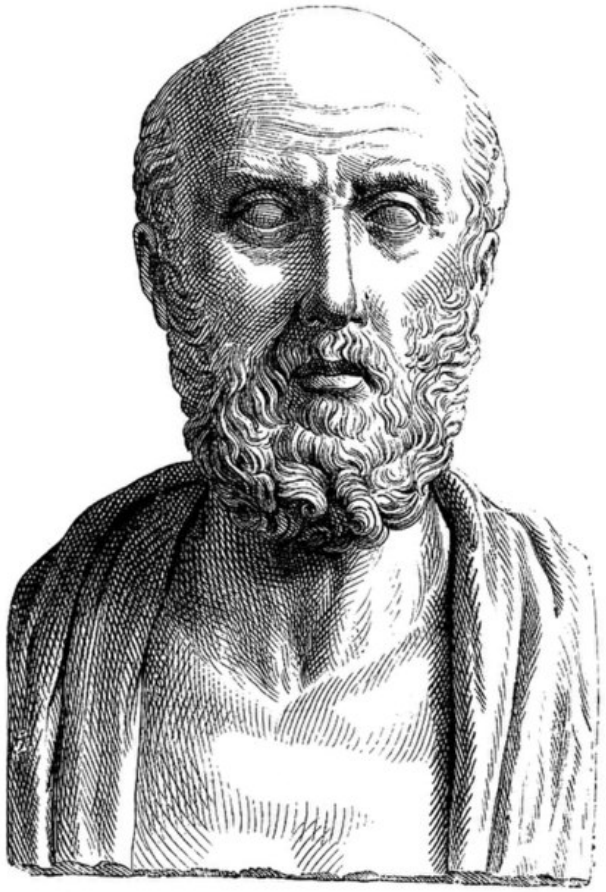
Prompt: a photo of a personal assistant;
Date: April 1, 2022

ON SYSTEMS & INCENTIVE STRUCTURES

- Novel systems are not, *de facto*, equalizers. They will tend toward propagating existing inequalities.
- Companies working on these systems may have conflicts of interest with respect to the incentive structures imposed 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



Professional ethics

Data Science Oath?

- First do no harm
- Preserve patient privacy
- Do not be afraid to say I don’t know
- Obligations to ALL human beings

BOX D.1
Hippocratic Oath

I swear to fulfill, to the best of my ability and judgment, this covenant:

I will respect the hard-won scientific gains of those physicians in whose steps I walk, and gladly share such knowledge as is mine with those who are to follow.

I will apply, for the benefit of the sick, all measures which are required, avoiding those twin traps of overtreatment and therapeutic nihilism.

I will remember that there is art to medicine as well as science, and that warmth, sympathy, and understanding may outweigh the surgeon's knife or the chemist's drug.

I will not be ashamed to say "I know not," nor will I fail to call in my colleagues when the skills of another are needed for a patient's recovery.

I will respect the privacy of my patients, for their problems are not disclosed to me that the world may know. Most especially must I tread with care in matters of life and death. If it is given me to save a life, all thanks. But it may also be within my power to take a life; this awesome responsibility must be faced with great humbleness and awareness of my own frailty. Above all, I must not play at God.

I will remember that I do not treat a fever chart, a cancerous growth, but a sick human being, whose illness may affect the person's family and economic stability. My responsibility includes these related problems, if I am to care adequately for the sick.

I will prevent disease whenever I can, for prevention is preferable to cure.

I will remember that I remain a member of society, with special obligations to all my fellow human beings, those sound of mind and body as well as the infirm.

If I do not violate this oath, may I enjoy life and art, respected while I live and remembered with affection thereafter. May I always act so as to preserve the finest traditions of my calling and may I long experience the joy of healing those who seek my help.

SOURCE: L.C. Lasagna, 1964, *Hippocratic Oath*, Modern Version, The Johns Hopkins Sheridan Libraries and University Museums. http://guides.library.jhu.edu/c.php?g=202502&p=1335759, accessed August 21, 2017.

BOX D.2
Data Science Oath

I swear to fulfill, to the best of my ability and judgment, this covenant:

I will respect the hard-won scientific gains of those data scientists in whose steps I walk and gladly share such knowledge as is mine with those who follow.

I will apply, for the benefit of society, all measures which are required, avoiding misrepresentations of data and analysis results.

I will remember that there is art to data science as well as science and that consistency, candor, and compassion should outweigh the algorithm's precision or the interventionist's influence.

I will not be ashamed to say, "I know not," nor will I fail to call in my colleagues when the skills of another are needed for solving a problem.

I will respect the privacy of my data subjects, for their data are not disclosed to me that the world may know, so I will tread with care in matters of privacy and security. If it is given to me to do good with my analyses, all thanks. But it may also be within my power to do harm, and this responsibility must be faced with humbleness and awareness of my own limitations.

I will remember that my data are not just numbers without meaning or context, but represent real people and situations, and that my work may lead to unintended societal consequences, such as inequality, poverty, and disparities due to algorithmic bias. My responsibility must consider potential consequences of my extraction of meaning from data and ensure my analyses help make better decisions.

I will perform personalization where appropriate, but I will always look for a path to fair treatment and nondiscrimination.

I will remember that I remain a member of society, with special obligations to all my fellow human beings, those who need help and those who don't.

If I do not violate this oath, may I enjoy vitality and virtuosity, respected for my contributions and remembered for my leadership thereafter. May I always act to preserve the finest traditions of my calling and may I long experience the joy of helping those who can benefit from my work.

National Academies of Science, Engineering, and Medicine. 2018. "Data Science for Undergraduates: Opportunities and Options." National Academies.

In the US, most students apply for grants or subsidized loans to finance their college education. Part of this process involves filling in a federal government form called the Free Application for Federal Student Aid (FAFSA). The form asks for information about family income and assets. The form also includes a place for listing the universities to which the information is to be sent. The data collected by FAFSA includes confidential financial information (listing the schools eligible to receive the information is effectively giving permission to share the data with them).

It turns out that the order in which the schools are listed carries important information. Students typically apply to several schools, but can attend only one of them. Until recently, admissions offices at some universities used the information as an important part of their models of whether an admitted student will accept admissions. The earlier in a list a school appears, the more likely the student is to attend that school.

Here's the catch from the student's point of view. Some institutions use statistical models to allocate grant aid (a scarce resource) where it is most likely to help ensure that a student enrolls. For these schools, the more likely a student is deemed to accept admissions, the lower the amount of grant aid they are likely to receive.

Is this ethical? Discuss.

Further resources

https://mdsr-book.github.io/mdsr2e/ch-ethics.html#professional-guidelines-for-ethical-conduct

For a book-length treatment of ethical issues in statistics, see Hubert and Wainer (2012). The National Academies report on data science for undergraduates National Academies of Science, Engineering, and Medicine (2018) included data ethics as a key component of data acumen. The report also included a draft oath for data scientists.

A historical perspective on the ASA's Ethical Guidelines for Statistical Practice can be found in Ellenberg (1983). The University of Michigan provides an EdX course on "Data Science Ethics." Carl Bergstrom and Jevin West developed a course Calling Bullshit: Data Reasoning in a Digital World". Course materials and related resources can be found at https://callingbullshit.org.

Andrew Gelman has written a column on ethics in statistics in *CHANCE* for the past several years (see, for example Andrew Gelman (2011); Andrew Gelman and Loken (2012); Andrew Gelman (2012); Andrew Gelman (2020)). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* describes a number of frightening misuses of big data and algorithms (O'Neil 2016).

The *Teach Data Science* blog has a series of entries focused on data ethics (https://teachdatascience.com). D'Ignazio and Klein (2020) provide a comprehensive introduction to data feminism (in contrast to data ethics). The ACM Conference on Fairness, Accountability, and Transparency (FAccT) provides a cross-disciplinary focus on data ethics issues (https://facctconference.org/2020).

The Center for Open Science—which develops the Open Science Framework (OSF)—is an organization that promotes openness, integrity, and reproducibility in scientific research. The OSF provides an online platform for researchers to publish their scientific projects. Emil Kirkegaard used OSF to publish his OkCupid data set.

The Institute for Quantitative Social Science at Harvard and the Berkeley Initiative for Transparency in the Social Sciences are two other organizations working to promote reproducibility in social science research. The *American Political Association* has incorporated the Data Access and Research Transparency (DA-RT) principles into its ethics guide. The Consolidated Standards of Reporting Trials (CONSORT) statement at (http://www.consort-statement.org) provides detailed guidance on the analysis and reporting of clinical trials.

Many more examples of how irreproducibility has led to scientific errors are available at http://retractionwatch.com/. For example, a study linking severe illness and divorce rates was retracted due to a coding mistake.