Data and Society
Data and Discrimination 2 – Lecture 9

March 1, 2021

Fran Berman, Data and Society, CSCI 4370/6370

Today's Class

- NEW ASSIGNMENT – **Op-ed** due 11:59 p.m. on March 15

- Lecture

- Student Presentations

Op-Ed Assignment and Instructions

- Write a **700-800 word Op-ed** on a **data-related topic** of *your choice*.
  - Choose something you feel strongly and persuade your reader about your point of view.

NYT Op-Ed: We don't need debates

FOX NEWS · FOX News Videos · September 29, 2020

Agweb Powered by Farm Journal

Op-Ed: Why Farmers Should Vote to Re-Elect Trump

Op-Ed by Charles Herbster, Chair of Farmers and Ranchers for Trump and owner and CEO of Conklin, Co. The opinions expressed are solely ...

3 hours ago

Agweb Powered by Farm Journal

Op-Ed: Why Farmers Should Elect Biden

Tom Vilsack served as the 40th governor of Iowa from 1999 to 2007, and as the 30th U.S. secretary of agriculture, under President Barack ...

3 hours ago

Why all New Yorkers, even the inherently suspicious, should download the new COVID alert app

The new New York, Covid alert app. As New York has attempted to safely re-open our businesses and vibrant cultural institutions, ...

11 mins ago

Hartford Courant

With everything else going wrong, the climate crisis is being downplayed. But it's the most important thing.

With everything else going wrong, the climate crisis is being downplayed. But it's the most important thing. By Mark A. Boyer.

4 days ago

Column: We have a fake eyelashes epidemic. I blame Fox News and the Kardashian sisters

Op-Eds are persuasive pieces

- Op-eds are **persuasive pieces**, so the **topic should be a little controversial**. (If everyone agrees on something, there is no reason to convince them).

    – Personal essays tell a personal story but are not necessarily trying to persuade.

- **Make sure your op-ed doesn't just *inform*** (provide information neutrally) **but *persuades*** (makes an evidence-based persuasive argument or advances an evidence-based opinion).

- Remember, **there needs to be a data angle** to your op-ed. (This is Data & Society …).

Largely uncontroversial topics	Good topics for op-eds
World peace is good.	We are heading towards a global war in the middle east.
Recycling is good for the environment.	We need to pass legislation to reduce the number of plastic bags.
Technology has transformed music.	Copyright laws are killing hip hop.

Op-Ed Detail -- Structure

Not all Op-Eds are like this, but many good Op-Eds have this structure:

- **Lede** – *Lead-in around a news hook or personal experience*

- **Thesis** – *your position (explicit or implied)*

- **Argument** – *should be based on evidence (stats, news, reports, expert quotes, scholarship, history, experience). Arguments often presented as a series of points.*

- **Criticism pre-emption** – *take the lead in acknowledging the flaws in your argument and address potential counter-arguments*

- **Conclusion** – *circle back to lede?*

Op-Ed Tips

- Write in a way that smart people can relate to, even if they are not in your discipline. **Don't use buzzwords** or talk "inside baseball" without explaining things.

- Pay attention to publication word count – op-eds are usually quite short

- *If you do this for real* (i.e. send it in to a publication rather than do it for class):
  - Timing is critical. A provocative thesis and lede can make a huge difference in terms of what gets published and when.
  - The final version may be reviewed and/or edited – what you send in may not be the final draft
  - Do your homework – everyone will read this. **Fact check, fact check, fact check**.
  - Be prepared for feedback – blogs, tweets, etc.
  - See http://www.nytimes.com/2013/10/14/opinion/op-ed-and-you.html?pagewanted=all&_r=0 for an article on writing an op-ed for the NY Times.

Grading Detail – Op-Ed   (15 points)

- Grade distribution:
  - **7 points on editorial content**: ideas, thesis, and supportive arguments
  - **8 points on writing**: does it work as an op-ed (addresses an issue, presents a well-evidenced opinion, is it more than an information piece …), does it tell a compelling story, is it well-written?

- **Submission: Send a .docX version to Fran by March 15, 11:59 p.m.**

- Op-eds should be in 12 pt. font, double-spaced and **between 700 and 800 words**. To support facts or cite information, **include references.**

Grading Rubric:

- **Content (7 points):**
  - Is the subject appropriate for an op-ed
  - Is the argument persuasive
  - Is the data angle clear?

- **Writing (8 points):**
  - Does the piece have all the elements of an op-ed?
  - Is the grammar, spelling, and flow appropriate for the piece?
  - Is the writing compelling?
  - Are there references as needed?

Reading for March 4

"**Scientists use big data to sway elections and predict riots – welcome to the 1960's**", Nature

https://www.nature.com/articles/d41586-020-02607-8

Date	Topic	Speaker	Date	Topic	Speaker
1-25	Introduction	Fran	1-28	The Data-driven World	Fran
2-1	Data and COVID-19	Fran	2-4	Data and Privacy -- Intro	Fran
2-8	Data and Privacy – Differential Privacy	Fran	2-11	Data and Privacy – Anonymity / Briefing Instructions	Fran
2-15	NO CLASS / PRESIDENT'S DAY		2-18	NO CLASS	
2-22	Legal Protections	Ben Wizner	2-25	Data and Discrimination 1	Fran
3-1	Data and Discrimination 2	Fran	3-4	Data and Elections 1	Fran
3-8	Data and Elections 2	Fran	3-11	NO CLASS / WRITING DAY	
3-15	Data and Astronomy (Op-Ed due)	Alyssa Goodman	3-18	Data Science	Fran
3-22	Digital Humanities	Brett Bobley	3-25	Data Stewardship and Preservation	Fran
3-29	Data and the IoT	Fran	4-1	Data and Smart Farms	Rich Wolski
4-5	Data and Self-Driving Cars	Fran	4-8	Data and Ethics 1	Fran
4-12	Data and Ethics 2	Fran	4-15	Cybersecurity	Bruce Schneier
4-19	Data and Dating	Fran	4-22	Data and Social Media	Fran
4-26	Tech in the News	Fran	4-29	Wrap-up / Discussion	Fran
5-3	NO CLASS				

Lecture 9 – Data and Discrimination 2

- **Facial Recognition and inaccuracy**
- **Joy Buolamwini (TED talk)**
- **Facial recognition as a tool for bias**

Facial recognition – protective and intrusive

- As facial recognition gets better, it is increasingly being used in a variety of environments – by police, animal conservationists, for public and private surveillance, etc. – and to ascertain a variety of characteristics.

- **Key questions for the public:**
    - How accurate are facial recognition algorithms?
    - How should we be using facial recognition algorithms?

Accuracy: Do face recognition algorithms represent race authentically?

- **National Institute of Standards and Technology 2019 Report:** How accurately do face recognition software tools identify people of varied sex, age and racial backgrounds?

- The answer depends on

    – What the training set is

    – What algorithm is being used

    – How the application uses it

- **Conclusion of the report: The majority of face recognition algorithms exhibit demographic differentials** (algorithm's ability to match two images of the same person varies from one demographic group to another).

Racial inaccuracy and Facial Recognition Algorithms

- **Joy Buolamwini, founder of the Algorithmic Justice League**

- **The Coded Gaze: Bias in Artificial Intelligence** (~13 min)

- https://www.youtube.com/watch?v=eRUEVYndh9c

Facial recognition as a tool for policing

Wrongfully Accused by an Algorithm

In what may be the first known case of its kind, a faulty facial recognition match led to a Michigan man's arrest for a crime he did not commit.

- Facial recognition match in Michigan led to arrest of Robert Julian-Borchak Williams for a crime he didn't commit.

- Photo for the match had come from 10 year-old driver's license photo and was not verified with other investigative methods before the arrest.

- Common problems with face matching algorithms include
    – Matching low resolution or partial images
    – Angle of surveillance camera not at eye level
    – User editing to normalize faces before searching
    – Biased training sets
    – Police using systems without training

Industry lightning rod: Clearview AI

- Clearview AI is a private tech company that provides facial recognition search system for police.

- Images uploaded to app which returns possible identity matches from Clearview DB (**3 billion pictures** scraped from public Facebook, YouTube, Venmo and millions of other websites)

- Clearview app has provided leads in many situations where none existed.

  – ~2400 law enforcement agencies use the service.

The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and "might lead to a dystopian future or something," a backer says.
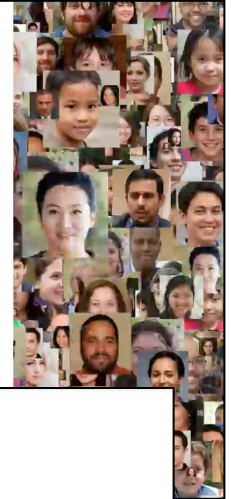
The facial-recognition app Clearview sees a spike in use after Capitol attack.

Jan. 9, 2021
By Kashmir Hill
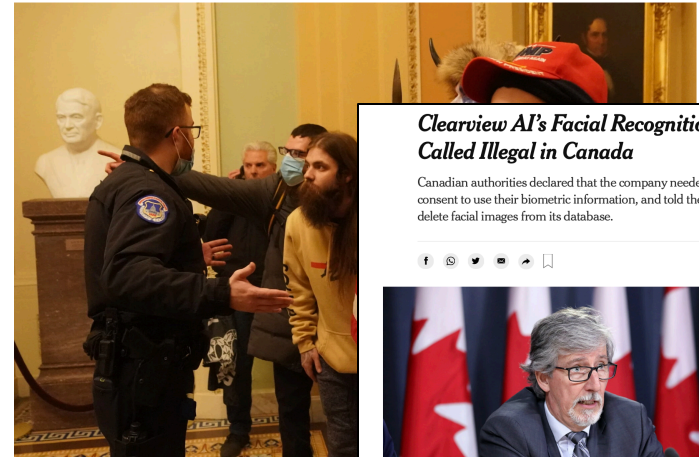
Trump supporters during the breach of the U.S. Capitol on Wednesday. Many of the rioters ... recognition technology. Erin Schaff/The New York Times

Clearview AI's Facial Recognition App Called Illegal in Canada

Canadian authorities declared that the company needed citizens' consent to use their biometric information, and told the firm to delete facial images from its database.

Canada's privacy commissioner, Daniel Therrien, said: "What Clearview does is mass surveillance, and it is illegal." Chris Wattie/Reuters

First Vermont data broker* enforcement action on Clearview

- State alleges that Clearview violated the Vermont data broker law by fraudulently acquiring brokered personal information through **screen scraping photographs** without consent and in violation of website terms of use, and **exposing sensitive personal data to theft** by foreign actors and criminals.
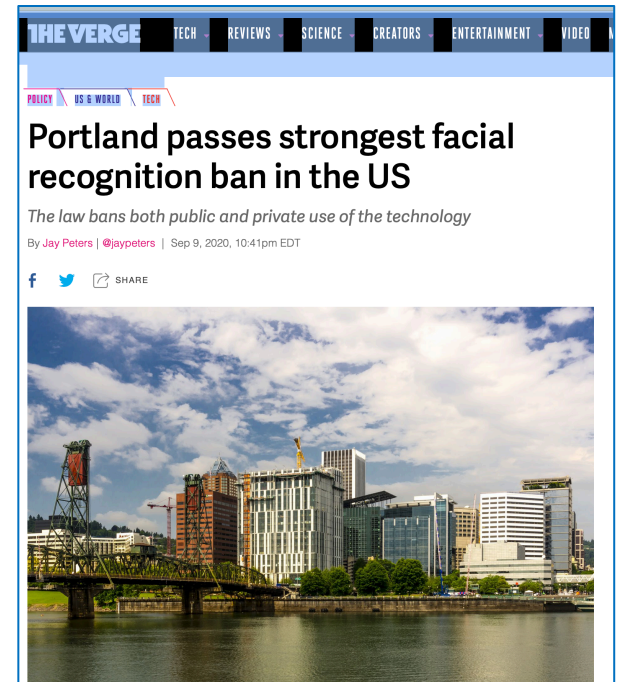
- State requests injunctive relief, restitution, disgorgement of Clearview's profits, and **penalties of $10,000 for each violation**.

- *Data broker = a business that "knowingly collects and sells or licenses to third parties the brokered personal information of a consumer with whom the business does not have a direct relationship."*

ReedSmith   Home

Technology Law Dispatch

HOME > PRIVACY & DATA PROTECTION > VERMONT ATTORNEY GENERAL BRINGS FIRST DATA BROKER ENFORCEMENT ACTION

Vermont Attorney General brings first data broker enforcement action

By *Divonne Smoyer, Samuel F. Cullari and Alexis Cocco* on 17 March 2020
Posted in *Privacy & Data Protection, Regulatory*

On March 10, 2020, Vermont Attorney General T.J. Donovan initiated an enforcement action based on Vermont's new data broker law against Clearview AI, Inc.

Vermont's data broker law, which became effective January 1, 2019, governs data brokers, which it defines as companies that collect and sell or license to third parties the personal information of a consumer with whom the business does not have a direct relationship. The law requires that data brokers (a) annually register with the Vermont Secretary of State, including completing certain necessary disclosures, and (b) maintain minimum data security standards. The law also prohibits any businesses or individuals – not just data brokers – from acquiring brokered personal information through fraudulent means or for the purpose of stalking, harassment, discrimination, or fraud.

According to the complaint, Clearview, which only registered as a Vermont data broker in January 2020 shortly before the publication of a New York Times article discussing many of the issues outlined in the complaint, uses "screen scraping" to amass a database of three billion photographs. Clearview then combines those photographs with facial recognition technology to create a commercial service that allows a customer to upload a photograph and "instantly identify the individual through facial recognition matching." While Clearview claims the technology exists to help law enforcement, the complaint alleges that Clearview has also provided its app to for-profit entities, investors, and foreign governments.

https://www.technologylawdispatch.com/2020/03/privacy-data-protection/vermont-attorney-general-brings-first-data-broker-enforcement-action/

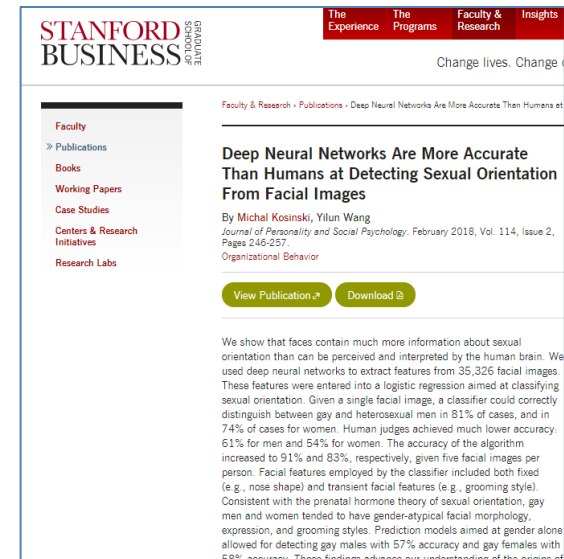Fran Berman, Data and Society, CSCI 4370/6370

Increasing state and federal concern about facial recognition

- Georgetown Law Center on Privacy and Technology conducted a survey that indicated that 1 in 2 US adults (117 million people) are in govt databases (compiled from driver's license photos and booking mugshots).

- **SF (CA), Oakland (CA), Somerville (MA), Brookline (MA) have passed bans on facial recognition use by public officials**

- CA banned facial recognition in police body cameras.

- Portland banned both public *and private* use of facial recognition

- **No current federal laws banning facial recognition.**

  - The ***Facial Recognition and Biometric Technology Moratorium*** is a current bill in Congress that would ban facial recognition and biometric surveillance technology by federal law enforcement agencies.

THE VERGE TECH REVIEWS SCIENCE CREATORS ENTERTAINMENT VIDEO

POLICY \ US & WORLD \ TECH

Portland passes strongest facial recognition ban in the US

The law bans both public and private use of the technology

By Jay Peters | @jaypeters | Sep 9, 2020, 10:41pm EDT

Facial recognition as a tool to ascertain sexual orientation -- The "Gaydar Study"

- **Science Problem:** Can a machine be used to detect sexual orientation, and how does it compare with human detection of sexual orientation?

- **Approach:** Deep neural network algorithm used to extract features from 35K facial images of gay and heterosexual men and women. Features processed and entered into a logistic regression to classify sexual orientation.

- **Results**: **Algorithm could correctly distinguish** between gay and heterosexual men in 81% of the cases and between gay and heterosexual women in 74% of the cases. **Human judges achieved lower accuracy**: 61% for men and 54% for women. (one facial image per person, algorithm accuracy increases with 5 facial images)

[Wang and Kosinski, 2017]

https://www.gsb.stanford.edu/faculty-research/publications/deep-neural-networks-are-more-accurate-humans-detecting-sexual

Fran Berman, Data and Society, CSCI 4370/6370

Theoretical basis for the study

- Approach based on **Prenatal Hormone Theory** (PHT) of sexual orientation: "same-gender orientation stems from the underexposure of male fetuses or the over-exposure of female fetuses to androgens that are responsible for sexual differentiation".

    – PHT predicts that gay men and women develop more gender atypical facial features than their heterosexual counterparts.

- Previous studies show mixed support for gender atypicality of facial features of gay men and women.

    – Previous studies used relatively small sample sizes
    – Also difficult to define what is "masculine" and what is "feminine"

Study Data Set

- **Data used from public images on on-line dating sites**
  - 36+K men/130+K images and 38+K women/170+K images used in study
  - Individuals classified heterosexual/homosexual dependent on "men seeking women/men", "women seeking men/women"
  - Respondents were not asked for consent other than through T&A on sites
  - Data from **largely white, U.S. constituency, ages 18-40**

- **Data needed to be normalized**: Images vary in quality, facial expression, head orientation, background, etc.
  - Facial recognition algorithm extracts key features
  - Derived dataset contained 50%/50% gay/straight men and 53%/47% gay/straight women

- **Subsequent data was also obtained from Facebook** websites popular among gay men according to "Facebook Audience Insights" platform and "interested in" field of user's FB profiles

Data preparation and normalization

Contour • Mouth • Eyebrows • Eyes • Nose

Pitch
Roll
Yaw
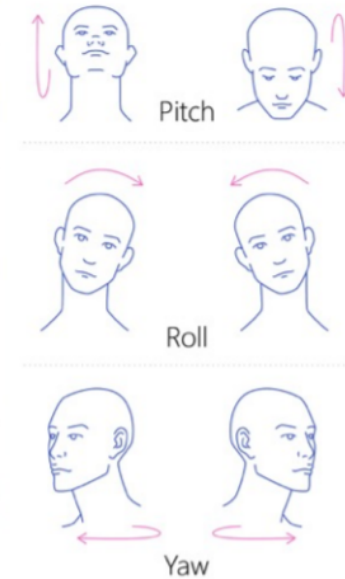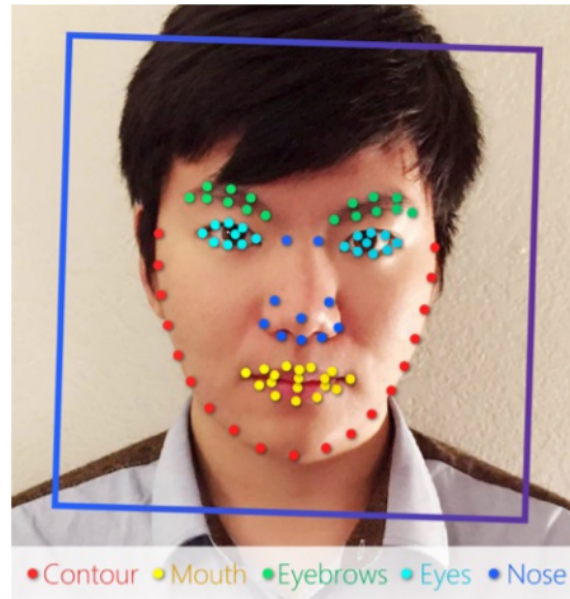
- Raw data taken from Facebook and dating sites (some faces had multiple images), run through a data preparation algorithm to normalize

  - Faces normalized using a set of "landmarks" to record the contour and features of the face as well as parameters providing the orientation of the face and head in space

- Scientists removed images from data set containing multiple faces, partially hidden faces, overly small faces, faces not facing the camera directly

- Amazon Turk workers verified that faces were adult, Caucasian, fully visible, of the gender reported

Is this good science?

- **Results (many and complex, see paper)**

  - **Algorithm could correctly distinguish** between gay and heterosexual individuals the majority of the time:
    - Gay and straight men in 81% of the cases
    - Gay and straight women in 74% of the cases.

  - **Human judges achieved lower accuracy:**
    - 61% for men
    - 54% for women

- **Limitations of the Study (per Authors):**

  - Images were non-standardized (varying quality, head orientation, facial expression)
  - Images obtained from a dating website – subjects may be posing in specific ways for the venue
  - Some data might be misleading (data not correct, user bisexual, etc.)
  - Insufficient number of non-white, non-US, gay subjects

> *"Our results provide strong support for the PHT, which argues that same-gender sexual orientation stems from the underexposure of male fetuses and overexposure of female fetuses to prenatal androgens responsible for the sexual differentiation of faces, preferences, and behavior."*
>
> Wang and Kosinski

Could facial recognition be used to discriminate based on sexual orientation?

- Authors on the study:

  - "Our findings suggest that publicly available data and conventional machine learning tools could be employed to build accurate sexual orientation classifiers."

  - "Additionally, given that **companies and governments are increasingly using computer vision algorithms to detect people's intimate traits**, our findings expose a threat to the privacy and safety of gay men and women."

  - "**We did not create a privacy-invading tool, but rather showed that basic and widely used methods pose serious privacy threats**."

Lecture 9 References (not already on slides)

- **Wrongfully accused by an algorithm,** New York Times, https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

- **Facing the algorithms**, World Magazine, https://world.wng.org/2021/02/facing_the_algorithms

- **Clearview,** Council on Foreign Relations, https://www.cfr.org/blog/canada-has-denounced-clearview-ai-its-time-united-states-follow-suit

- **The secretive company that might end privacy as we know it**, New York Times, https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

- **Vermont data broker law**, Lexology, https://www.lexology.com/library/detail.aspx?g=5398e04c-f672-428e-b8f6-707538072acf

- **Gaydar study**, https://www.gsb.stanford.edu/sites/gsb/files/publication-pdf/wang_kosinski.pdf

Presentations

Upcoming Presentations

March 4

- "**Estonia leads the world in making digital voting a reality**", Financial Times, https://www.ft.com/content/b4425338-6207-49a0-bbfb-6ae5460fc1c1

- "**Experts issue warning about electronic voting; blockchain-based systems are not ready for prime time**", CPO Magazine, https://www.cpomagazine.com/cyber-security/mit-security-experts-issue-warning-about-electronic-voting-blockchain-based-systems-are-not-ready-for-prime-time/

March 8

- "**Election forecast models are worth more attention than polls**", Bloomberg Opinion, https://www.bloomberg.com/opinion/articles/2020-11-22/election-forecast-models-have-more-potential-than-simple-polling (Chris P.)

- "**Which 2020 election polls were most – and least – accurate?**", Washington Post, https://www.washingtonpost.com/politics/2020/11/25/which-2020-election-polls-were-most-least-accurate/ (Isaac L.)

Need Volunteers

March 15

- "**Hunting for a giant black hole, astronomers found a nest of darkness**," New York Times, https://www.nytimes.com/2021/02/26/science/astronomy-black-hole-ngc6397.html (Greg S.)

- "**NASA's Webb telescope will be the world's premier space science observatory – Here's what those those powerful capabilities mean for astronomy**," SciTechDaily, https://scitechdaily.com/nasas-webb-telescope-will-be-the-worlds-premier-space-science-observatory-heres-what-those-powerful-capabilities-mean-for-astronomy/ (Angelina M.)

Presentations for Today

March 1

- "**Is an Algorithm less Racist than a Loan Officer?**", New York Times, https://www.nytimes.com/2020/09/18/business/digital-mortgages.html

- "**Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use**", Washington Post, https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/