Ethics, Public Policy, and Technological Change

Rob Reich
Mehran Sahami
Head TA: Roberta Fischli

Housekeeping

- Assignment #1 grades released
  - Regrades open for one week
- Assignment #2 (Philosophy Paper) due February 12
  - Sample philosophy papers available on website
  - WIM (CS182W) students only: sign up for a slot with a TCP writing tutor as soon as possible. Link on class website.
- Feel free to come down to the tech-free zone in the front row

Today's Agenda

1. What can we infer from your digital trails?
2. The information ecosystem
3. Other technical considerations in data privacy
   - Data storage and encryption
   - Multi-party privacy
4. Facial Recognition
5. Policy considerations

Today's Agenda

1. **What can we infer from your digital trails?**
2. The information ecosystem
3. Other technical considerations in data privacy
   - Data storage and encryption
   - Multi-party privacy
4. Facial Recognition
5. Policy considerations

But I Just Pressed the Like Button

- Kosinski *et al* (2013) analyze "Likes" on Facebook to see what they can infer about you
  - 58,466 volunteers provided "Likes", demographics, and psychometrics
  - Average of 170 "Likes" per person
  - Built predictive models for various attributes based only on "Likes"
  - Even one "Like" carries predictive information!

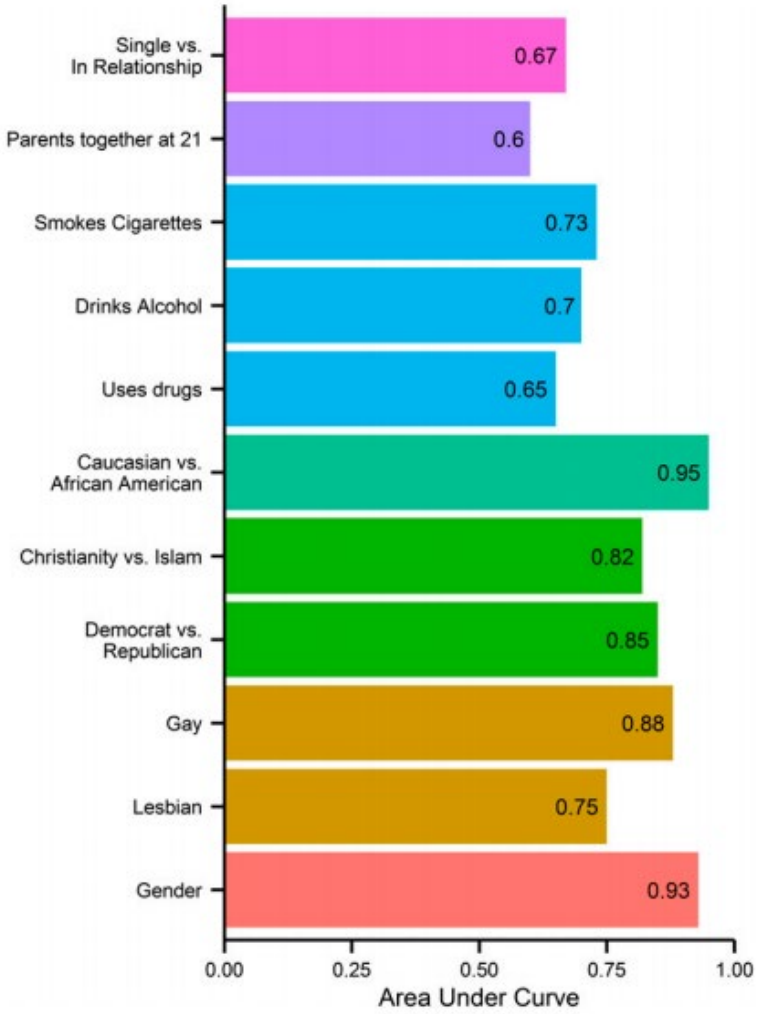


Fig. 2. Prediction accuracy of classification for dichotomous/dichotomized attributes expressed by the AUC.

Problematic Work

- In 2017, Wang and Kosinski publish "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images"
  - Analyze 35,326 facial images
  - *"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. Human judges achieved much lower accuracy: 61% for men and 54% for women."*
- BBC News: "Two US-based LGBT-focused civil rights groups issued a joint press release attacking the study in harsh terms."
  - *"These reckless findings could serve as a weapon to harm both heterosexuals who are inaccurately outed, as well as gay and lesbian people who are in situations where coming out is dangerous." [said Jim Halloran, chief digital officer of Glaad, a media-monitoring body.]*

Source: Row over AI that 'identifies gay faces', BBC News, 11 September 2017, https://www.bbc.co.uk/news/technology-41188560.amp

Recommendation Systems

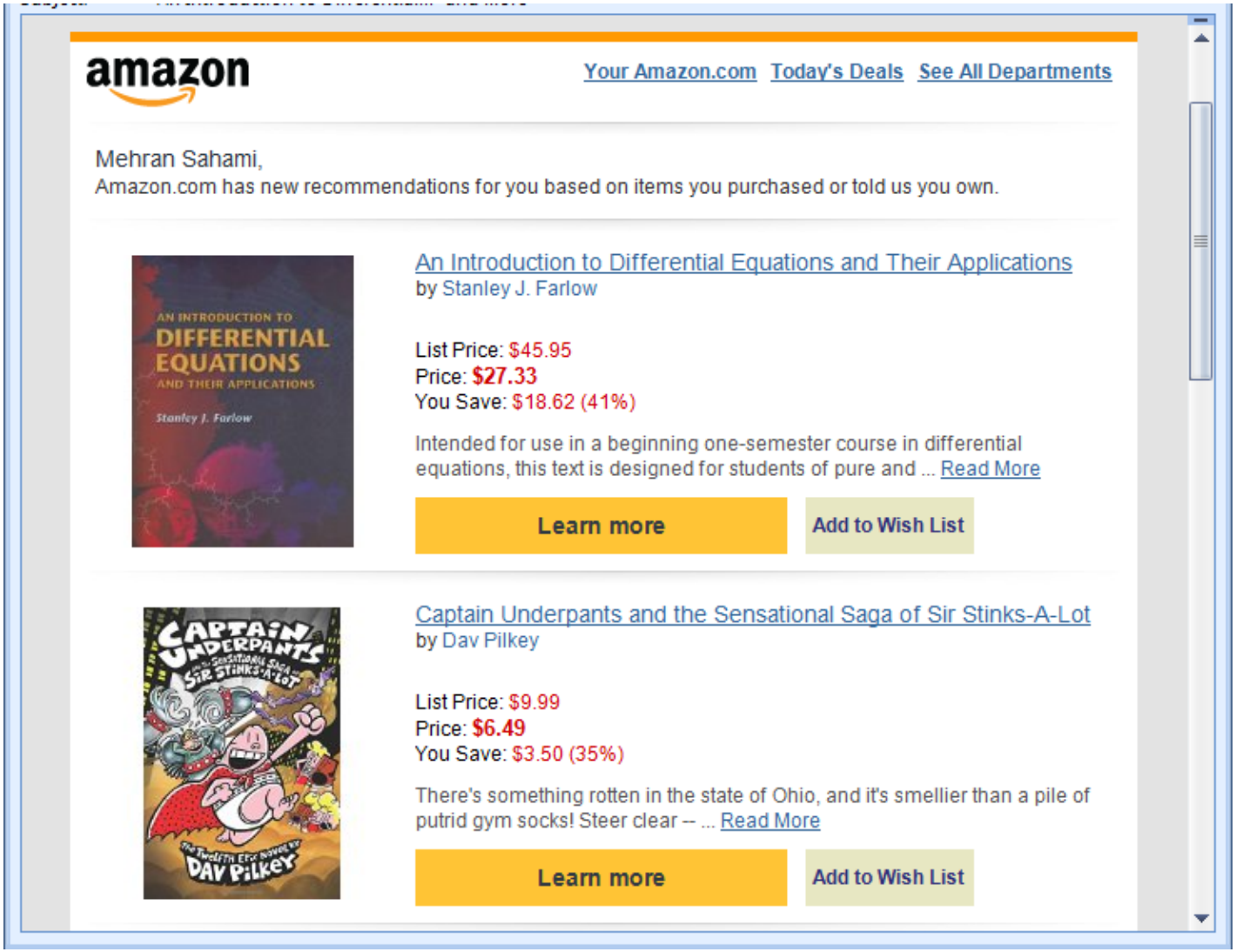


Image: Screenshot of personal email from Amazon.com to Mehran Sahami

Today's Agenda

1. What can we infer from your digital trails?
2. **The information ecosystem**
3. Other technical considerations in data privacy
   - Data storage and encryption
   - Multi-party privacy
4. Facial Recognition
5. Policy considerations

An Information Ecosystem

- Anyone heard of LiveRamp/Acxiom?
  - $1.5B company headquartered in San Francisco
  - Provide data for marketing segmentation, which are even named
  - Let's look at their "Life Stage Clustering System: 'Personicx'"

- "Full Steaming" is a mix of affluent, well-educated couples and singles that have a net worth exceeding $500,000…

- "Lavish Lifestyles" contains established couples with teenage kids, minivans and mortgages…

- "Tots & Toys" is dominated by affluent and well-educated working couples with preschool-age children. They are homeowners, mainly in single-family houses…

…Where Everyone is Classified…

- "Truckin' & Stylin'" households are in their mid- to late-30s and live in rural towns. They rank just below average for household income but drop to the bottom of the list (66th) for net worth…

- "Resilient Renters" represents an ethnically diverse group of singles. They are renters and, if employed, earn extremely low wages in clerical and blue-collar jobs. This cluster represents one of the lowest for net worth…

- At a mean age of 25, "Early Parents" represents one of the youngest of the segments. It contains single and married parents in their mid-20s whose spending habits and leisure time are heavily influenced by their young children. Early Parents ranks among the nation's lowest clusters for income and net worth…

…Including You, “Collegiate Crowd”

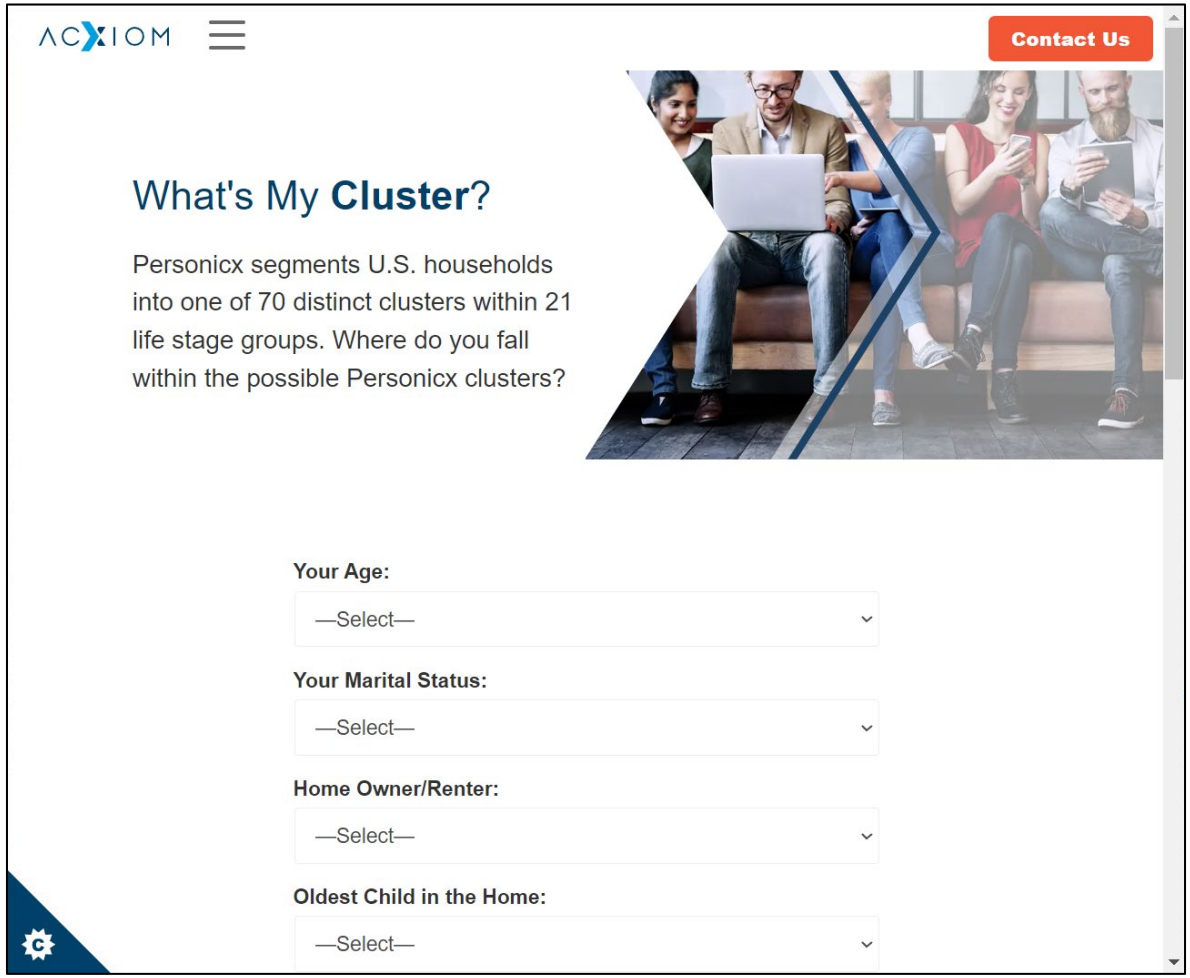
With a mean age of 21, this group represents the youngest of all the clusters. The cluster has a high concentration of students, a correlating low income and net worth, and high mobility.

Collegiate Crowd is made up of single, highly mobile apartment dwellers. Over half of all Collegiate Crowd live in the big-three Central regions, 82% in metropolitan areas boasting colleges and universities. The group has a 31% adult concentration of students, and ranks 53rd for household income and 50th for net worth.

Not surprisingly, they are twice as likely to have education loans. This group is constantly online using laptops and the Internet for news, weather and job hunting. As for activities, volleyball, basketball, barhopping and movie going fill their time, and this group is much more likely than average to attend college football games.

You Can Check Your Cluster

https://www.acxiom.com/whats-my-cluster/



The Information Ecosystem

- Are there any benefits here?

- Consider the example Rob gave you of Target building a predictive model to identify pregnant women
  - What if this was US government trying to reach women to provide:
    - Information about and access to prenatal health care
    - Nutritional information and supplemental nutrition assistance

  - Also allows for longer-term tracking
    - When child is toddler (pre-K education)
    - When child goes to school (school options/vouchers)
    - When child might prepare to go to college (financial aid)
    - When child becomes an adult

Today's Agenda

1. What can we infer from your digital trails?
2. The information ecosystem
3. **Other technical considerations in data privacy**
   - Data storage and encryption
   - Multi-party privacy
4. Facial Recognition
5. Policy considerations

Data Storage

- Recall that data privacy involves competing interests
- A conversation from a few years ago:
  - Grade data should not be stored on non-Stanford systems
  - E.g., Should not use Google spreadsheet to store students grades
  - Reason: security of the underlying data (FERPA)
- *Verifiably true story time!*
  - The story of the missing PhD thesis
- Are you concerned about your grades in this class being stored in a Google spreadsheet? If so, what might mitigate your concerns?

Encryption

- Perhaps we could encrypt all stored data
  - If there is a system breach, data is “safe”
  - Encryption can have vulnerabilities in practice
    - Data is often decrypted for analysis (and left that way)
    - Breach could lead to leak of decryption key
- Perhaps government should be able to decrypt data for security
  - Encryption systems could provide a backdoor
  - E.g., NSA can have access to “key escrow” to get decryption key
- This has been a debate for decades
  - E.g., Clipper Chip in 1990s

Content Scanning

- In 2021, Apple announced it would scan photos on iCloud for child sexual abuse material (CSAM)
  - Apple ***stops release*** of this feature after public outcry
  - Objection: such scanning could be extended by different governments interested in finding other content they find objectionable
- Proposals also exist for in-app content moderation in end-to-end encrypted messaging systems
  - Machine learning model in messaging app that detects harmful content
  - E.g., Apple's Communication Safety in Messages feature

The operating system analyzes image and video attachments and determines if the content contains nudity without sending information off the device. The feature is designed so that no indication of the detection of nudity leaves the device. Apple does not get access to the messages, and no notifications are sent to the parent or anyone else.

Source: https://www.apple.com/child-safety/, Accessed February 8, 2026.

Multi-Party Privacy

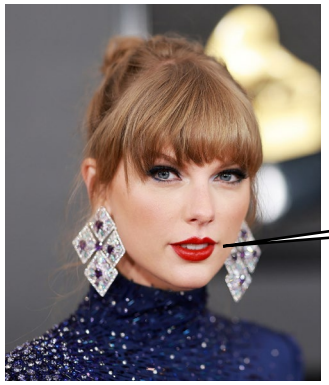
- Traditional privacy set-up with a data subject and a data user
  - Platform wants to use your data for personalization and marketing
  - Medical researcher want to use your data to improve healthcare
  - NSA needs to determine contents of phone for national security
- Increasingly live in multi-party privacy settings
  - Alice posts a picture (or a tweet) tagging Bob and Chris
  - Bob may untag himself (if he is notified about tag)
    - Untagging may offend Alice, so Bob may not do it (modify behavior)
    - Bob's friends may still see post (if friends with Alice or Chris)
    - Untagging or even deleting post may not matter, if post was copied
  - Bob can (and, research shows, will) change behavior with respect to Alice to avoid being in posts (i.e., unwanted privacy violations)
- Have you modified your behavior with others to prevent multi-party privacy breaches? Or worry about offending others?

Today's Agenda

1. What can we infer from your digital trails?
2. The information ecosystem
3. Other technical considerations in data privacy
   - Data storage and encryption
   - Multi-party privacy
4. **Facial Recognition**
5. Policy considerations

Facial Recognition

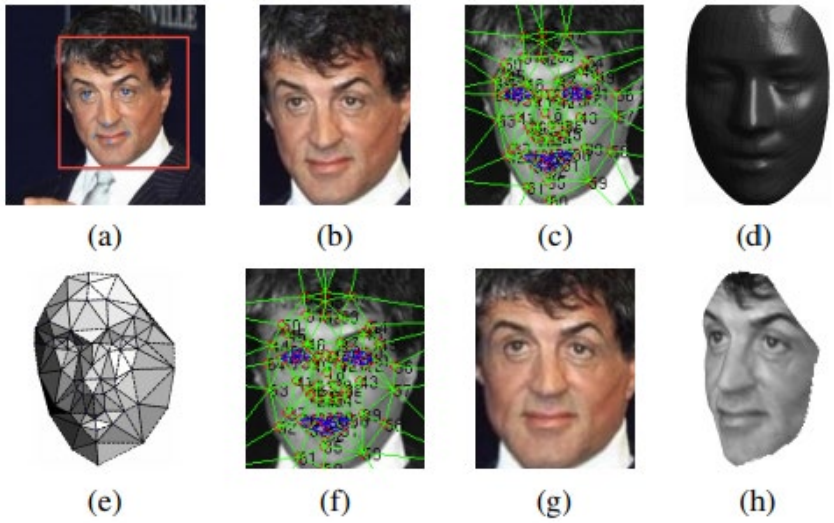
- Applications
  - Verification (e.g., unlocking you iPhone with FaceID)
  - Tagging (e.g., Facebook photos)
    - Note: After user pushback, Facebook/Meta no longer tags photos
  - Many others discussed in case study in sections
- Surveillance
  - Chinese police use facial recognition at concert with 60,000 people to identify and arrest suspect
  - Facial recognition used at Taylor Swift concerts to identify stalkers

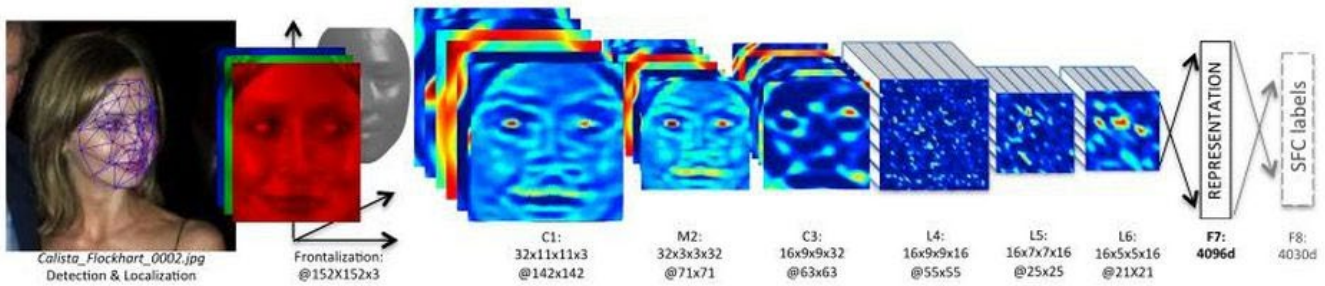


The problem is more than just one person!

Meta/Facebook DeepFace

- Facebook DeepFace algorithm (Taigman *et al*, 2014)
  - Nine layer neural network with 120+ million weights
  - Trained on 4+ million images
  - Build 3D model from 2D image to get “frontal” view
  - Claim 97.35% accuracy
  - Comparable to humans





Images from: Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1701-1708, doi: 10.1109/CVPR.2014.220.

Issues with Facial Recognition

- Facial recognition suffers from the same biases as other forms of algorithmic decision-making
  - Gender Shades project by Joy Buolamwini and Timnit Gebru
  - Video

- Issue of classification accuracy
  - Less data for darker-skinned people, especially women
- But, in criminal facial recognition, often have more data of darker-skinned people

Images (Top to Bottom): Niccolò Caranti , Wikimedia (CC-BY-SA 4.0); TechCrunch, Wikimedia (CC-BY 2.0)

Classification vs. Matching

- In classification task, we are learning parameters for model based on data
  - This is the case when we are determining *if* an image contains a face
  - ***More data often means better classification***
  - Built in assumption that the testing (actual usage) data is same distribution as training data
  - So, if we don't do a good job classifying groups with limited data in training set, we don't expect to get penalized for this in the testing set (i.e., actual usage)
- Recall the classifier that is 99.5% correct if it predicts "negative" all the time and only 0.5% of population are positive for some condition

Classification vs. Matching

- In *matching* task, we are trying to find closest matching image in a dataset
  - This is the case when we have image of a face and are trying to *identify* whose face it is
  - ***More data often means higher likelihood of a match***
  - Go ahead, upload more pictures of yourself to the web. I'll wait.
- In criminal justice context, higher likelihood of match can mean higher chance of misidentification
- African-Americans often under-represented in datasets for *classification* of faces and can be over-represented in datasets for *matching* of faces
  - They are hurt in both the classification and matching contexts

Misidentification

Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots

By Jacob Snow, Technology & Civil Liberties Attorney, ACLU of Northern California
JULY 26, 2018 | 8:00 AM

"Nearly 40 percent of Rekognition's false matches in our test were of people of color, even though they make up only 20 percent of Congress."

"In a recent letter to Amazon CEO Jeff Bezos, the Congressional Black Caucus expressed concern about the 'profound negative unintended consequences' face surveillance could have for Black people, undocumented immigrants, and protesters."

Image: ACLU, "Amazon's Face Recognition Falsely Matched 28 Members of Congress With Mugshots" (July, 2018), https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28

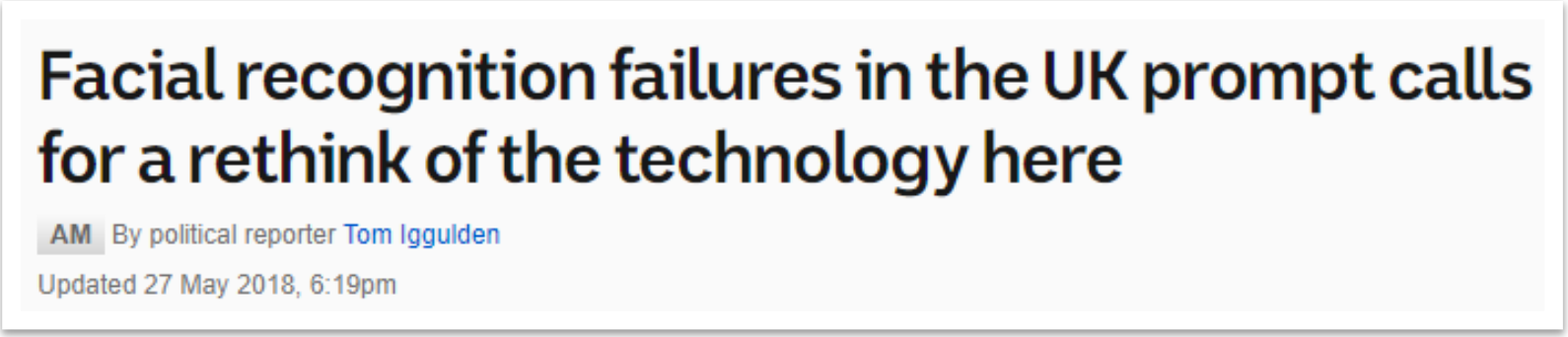
London

It is estimated that there are 500,000 CCTV cameras dotted around London.

The average person living in London will be recorded on camera 300 times in one day

https://www.caughtoncamera.net/news/how-many-cctv-cameras-in-london/

Let's Rethink This



Facial recognition failures in the UK prompt calls for a rethink of the technology here

AM By political reporter Tom Iggulden

Updated 27 May 2018, 6:19pm

"London police trial of facial recognition technology generated 104 'alerts', of which 102 were false."

"Another trial by South Wales police returned 2,400 false positives from CCTV footage gathered at UEFA football matches and the like."

"Legislation currently before the Australian Parliament would allow national security agencies to use driver's license photos and, potentially, social media images to match with CCTV footage."

Okay, now go delete all those images from the web. I'll wait.

Oh, Let's Use it Anyway

TIME SUBSCRIBE

BY BILLY PERRIGO JANUARY 24, 2020

London Police to Deploy Facial Recognition Cameras Despite Privacy Concerns and Evidence of High Failure Rate

"The Metropolitan Police, the U.K.'s biggest police department with jurisdiction over most of London, announced Friday it would begin rolling out new 'live facial recognition' cameras in London, making the capital one of the largest cities in the West to adopt the controversial technology."

"Privacy activists immediately raised concerns, noting that independent reviews of trials of the technology showed a failure rate of 81%."

Source: Billy Perrigo (2020), "London Police to Deploy Facial Recognition Cameras Despite Privacy Concerns and Evidence of High Failure Rate," Time, https://time.com/5770976/london-facial-recognition-police/

Punitive Uses of Facial Recognition

The New York Times

By Kashmir Hill and Corey Kilgannon

Published Dec. 22, 2022 Updated Jan. 3, 2023

Madison Square Garden Uses Facial Recognition to Ban Its Owner's Enemies

MSG Entertainment, the owner of the arena and Radio City Music Hall, has put lawyers who represent people suing it on an "exclusion list" to keep them out of concerts and sporting events.

Over Thanksgiving weekend, Kelly Conlon, 44, a personal injury lawyer from Bergen County, N.J., was chaperoning her 9-year-old daughter's Girl Scout troop on a trip … [to] Radio City Music Hall.

Before she could even glimpse the Rockettes, however, security guards pulled Ms. Conlon aside and her New York jaunt took an Orwellian turn.

"They told me that they knew I was Kelly Conlon and that I was an attorney," she said this week. "They knew the name of my law firm."

The guards had identified her using a facial recognition system.

Sources: Kashmir Hill and Corey Kilgannon, "Madison Square Garden Uses Facial Recognition to Ban Its Owner's Enemies," New York Times, Published Dec 22., 2022, Updated Jan. 3, 2023. https://www.nytimes.com/2022/12/22/nyregion/madison-square-garden-facial-recognition.html

San Francisco in 2019

The New York Times

San Francisco Bans Facial Recognition Technology

By Kate Conger, Richard Fausset and Serge F. Kovaleski
May 14, 2019

SAN FRANCISCO — San Francisco, long at the heart of the technology revolution, took a stand against potential abuse on Tuesday by banning the use of facial recognition software by the police and other agencies.

The action, which came in an 8-to-1 vote by the Board of Supervisors, makes San Francisco the first major American city to block a tool that many police forces are turning to in the search for both small-time criminal suspects and perpetrators of mass carnage.

Image: Screenshot https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html#:~:text=San%20Francisco%20banned%20the%20use,on%20the%20streets%2C%20in%20parks.

And San Francisco in 2024

WIRED

NEWSLETTERS SUBSCRIBE

LAUREN GOODE TOM SIMONITE BUSINESS MAR 6, 2024 6:05 PM

5 Years After San Francisco Banned Face Recognition, Voters Ask for More Surveillance

On Tuesday the country's techiest city backed a ballot proposition that tapped into concerns about crime, giving the police more freedom to use drones and other surveillance technology.

Proposition E passed with 60 percent of the vote and was backed by San Francisco mayor London Breed. It gives the San Francisco Police Department new freedom to install public security cameras and deploy drones without oversight from the city's Police Commission or Board of Supervisors.

Image: https://www.wired.com/story/san-francisco-banned-face-recognition-voters-ask-for-more-surveillance/

Government Surveillance Today

The Washington Post
Democracy Dies in Darkness

The powerful tools in ICE's arsenal to track suspects — and protesters

Biometric trackers, cellphone location databases and drones are among the surveillance technologies that federal agents are tapping in their deportation campaign.

By Eva Dou, Artur Galocha and Kevin Schaul

January 29, 2026

ICE amps up its surveillance powers, targeting immigrants and antifa

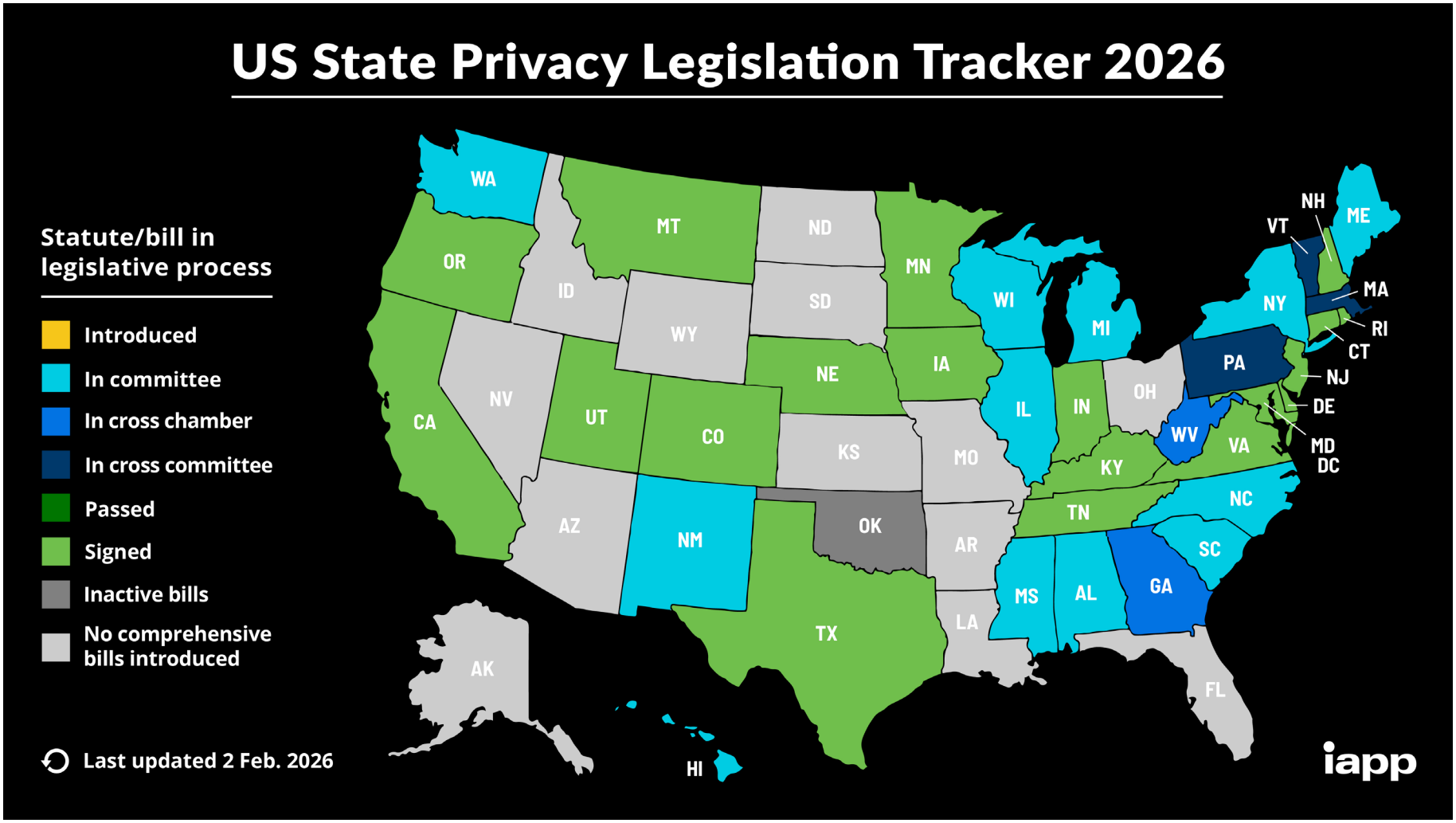
Iris scanners, facial-recognition apps, phone-hacking software and cellphone location data are among the agency's recent technological purchases.

Updated October 17, 2025

Today's Agenda

1. What can we infer from your digital trails?
2. The information ecosystem
3. Other technical considerations in data privacy
   - Data storage and encryption
   - Multi-party privacy
4. Facial Recognition
5. **Policy considerations**

Privacy Legislation is a Patchwork



GDPR vs. CCPA/CPRA

	GDPR	CCPA/CPRA
Right to know	✓	✓
Right to access	✓	✓
Right to forget	✓	✓
Right to data portability	✓	✓
Right to correct	✓	✓

GDPR vs. CCPA/CPRA

	GDPR	CCPA/CPRA
Right to know	✓	✓
Right to access	✓	✓
Right to forget	✓	✓
Right to data portability	✓	✓
Right to correct	✓	✓
Consent	Opt in	Opt out (unless < 16)
Nondiscrimination		✓

GDPR vs. CCPA/CPRA

	GDPR	CCPA/CPRA
Right to know	✓	✓
Right to access	✓	✓
Right to forget	✓	✓
Right to data portability	✓	✓
Right to correct	✓	✓
Consent	Opt in	Opt out (unless < 16)
Nondiscrimination		✓
Covered entities	Data controllers	$25M revenue > 100k households
Enforcement	Private and public	Private only for data breaches

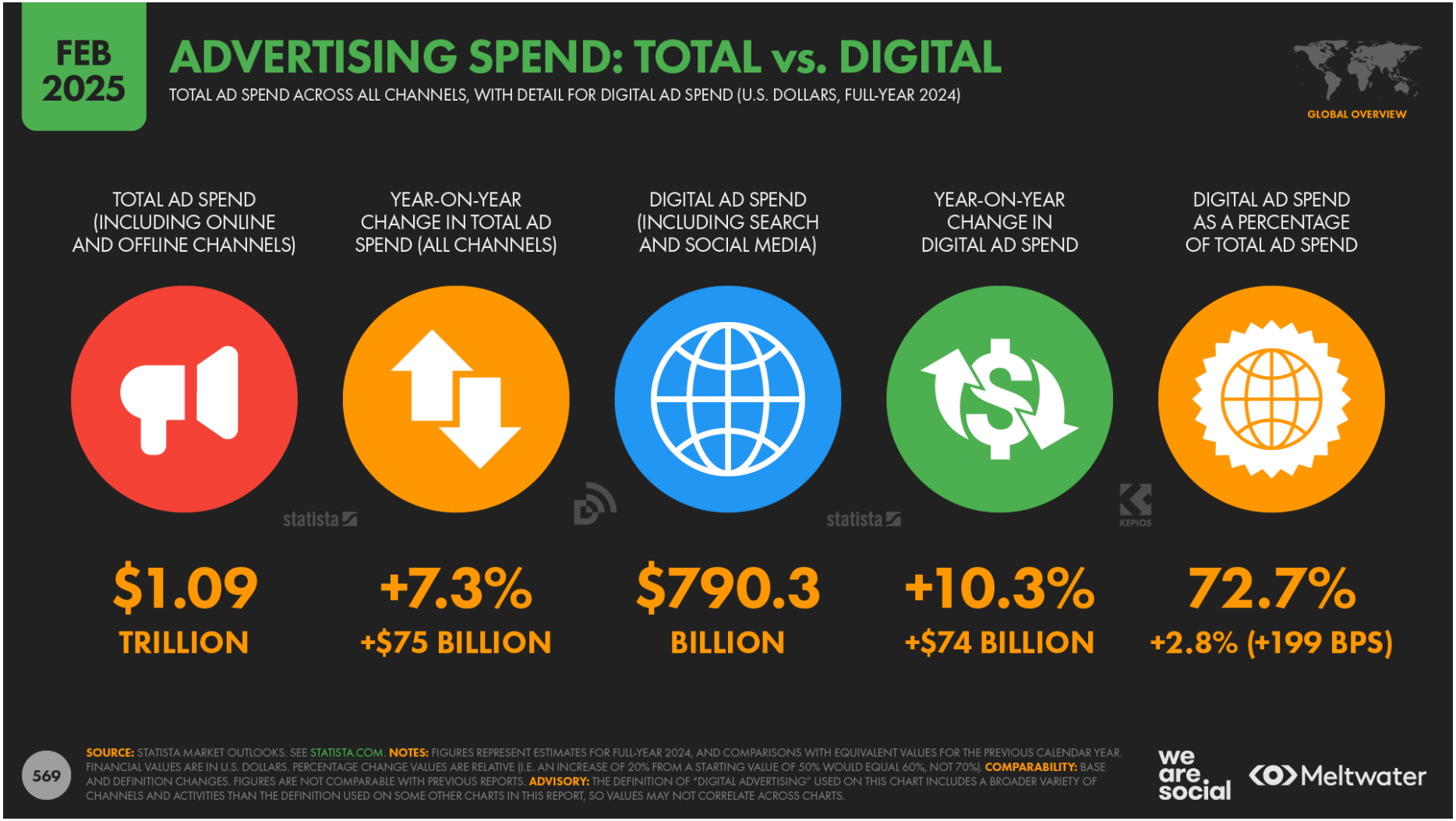
EU vs. US (CA)

Why does Europe adopt more a more assertive regulatory approach, while the United States relies on self-regulation and privacy self-management?

Reasons for differences between GDPR and CCPA/CPRA

- Different values (as a function of experience, constitution, etc.)
  - Framework: Human rights vs. Harms
- Pressure from organized interests, groups
- Partisan and electoral competition
- Bureaucratic interests

There is a Lot at Stake



Privacy as Competitive Advantage

Image: Apple advertisement

Alternative Business Models

neeva

Neeva was founded in 2019 by Sridhar Ramaswamy and Vivek Raghunathan. Neeva had raised $77.5m from investors.

In 2023, the co-founders announced the shutdown of the search engine on June 2.

According to them, the main reason for that decision was how hard it was to persuade normal users to make the switch.

Source: www.neeva.com; Wikimedia Commons CC BY-SA 4.0 and https://en.wikipedia.org/wiki/Neeva