Today’s Agenda

1. Technical considerations in data privacy
   • Data storage
   • Multi-party privacy
   • Information discovery
   • Issues in facial recognition
2. Regulatory considerations in data privacy
3. Philosophical considerations in data privacy
Today’s Agenda

1. Technical considerations in data privacy
   • Data storage
   • Multi-party privacy
   • Information discovery
   • Issues in facial recognition
2. Regulatory considerations in data privacy
3. Philosophical considerations in data privacy
Data Storage

• Recall that data privacy involves competing interests

• 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!*
  • Where did my PhD thesis go?

• Are you concerned about your grades in this class being stored in a Google spreadsheet? If so, what might mitigate your concerns?
Today’s Agenda

1. Technical considerations in data privacy
   • Data storage
   • **Multi-party privacy**
   • Information discovery
   • Issues in facial recognition

2. Regulatory considerations in data privacy

3. Philosophical considerations in data privacy
Multi-Party Privacy

• Traditional privacy set-up with a data subject and a data user
  • Platform wants to use you 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?
1. Technical considerations in data privacy
   • Data storage
   • Multi-party privacy
   • Information discovery
   • Issues in facial recognition
2. Regulatory considerations in data privacy
3. Philosophical considerations in data privacy
Information Discovery

• Data minimalism
  • Only collect the data needed at time of data gathering
  • View espoused by GDPR

• Data maximalism
  • Gather as much data as possible
  • Don’t know a priori what data may be useful later
  • View often espoused by start-ups

• Is choice context dependent?
  • What would you want Google to do?
  • What would you want your medical practitioner to do?
1. Technical considerations in data privacy
   • Data storage
   • Multi-party privacy
   • Information discovery
   • Issues in facial recognition
2. Regulatory considerations in data privacy
3. Philosophical considerations in data privacy
Facial recognition suffers from the same biases as other forms of algorithmic decision-making
- Gender Shades project by Joy Buolamwini (Masters thesis)
- Initial work done jointly with 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
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 data seen in deployment is same distribution as data used to train/validate the model
  - So, if we don’t do a good job classifying groups with limited data in development, we don’t expect a large penalty in 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

- What are approaches to mitigate this problem?
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 Facebook. I’ll wait.

• In criminal justice context, higher likelihood of match can mean higher chance of misidentification as a potential criminal

• 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
“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.”
“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 Facebook images. I’ll wait.
“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%.”
A Sober Assessment

• There are deep privacy issues with data collection/aggregation and applications based on it (e.g., facial recognition)
  • Neither despair nor resignation is helpful; neither is hyperbole
  • We may choose to sacrifice privacy for things we want, for example:
    • Greater security
    • Transactional convenience (easy communication, free access to information/search, social connection, consumer efficiency, etc.)
    • Improved social services (healthcare, transportation, etc.)
  • Need to understand there are distributional effects

• Grappling with tough questions
  • What are appropriate privacy guarantees to demand from the private and public sector?
  • What should consumers be expected to do (if anything)?
    • How can they get the information needed to make choices?
  • How much of this should be regulated by government?
Today’s Agenda

1. Technical considerations in data privacy
   • Data storage
   • Multi-party privacy
   • Information discovery
   • Issues in facial recognition

2. **Regulatory considerations in data privacy**

3. Philosophical considerations in data privacy
### Three Models of Regulation

<table>
<thead>
<tr>
<th><strong>Self-Regulation</strong></th>
<th><strong>Privacy Self-Management</strong></th>
<th><strong>Regulation by Supervisory Bodies</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Information practices are implemented internally by organizations with limited external intervention</td>
<td>• Control is ensured through the active participation and intervention of those whose data is held</td>
<td>• Interaction between data subjects and controllers is mediated by a third party</td>
</tr>
</tbody>
</table>
Privacy Self-Management

Can it work?
1. Transparency (terms of service/cookies)
2. Warning labels (be like the FDA!)
3. Distinct, competing products (Apple vs. alternatives)
4. Opt-in vs. opt-out (more paternalistic)

Should we rely on individuals to make these choices?
1. Yes, freedom!
2. No, it’s just...
   • Too much information
   • Too hard
   • Too costly to society
Industry Self-Regulation

**GENERAL AUDIENCES**

- **G**: All Ages Admitted

  - Nothing that would offend parents for viewing by children.

**PARENTAL GUIDANCE SUGGESTED**

- **PG**: Some Material May Not Be Suitable for Children

  - Parents urged to give "parental guidance." May contain some material parents might not like for their young children.

**PARENTS STRONGLY CAUTIONED**

- **PG-13**: Some Material May Be Inappropriate for Children Under 13

  - Parents are urged to be cautious. Some material may be inappropriate for pre-teenagers.

**RESTRICTED**

- **R**: Under 17 Requires Accompanying Parent or Adult Guardian

  - Contains some adult material. Parents are urged to learn more about the film before taking their young children with them.

**NO ONE 17 AND UNDER ADMITTED**

- **NC-17**: No One 17 and Under Admitted

  - Clearly adult. Children are not admitted.
How confident are you in relying on companies to...

- To regulate themselves?
- To self-regulate as an industry?
- To be responsive to user pressure?
- To be responsive to their own employees?

**How dominant?**

Global market share
April 2018, %

**Search**
- Google 91%

**Smartphone web traffic**
- Apple 45%

**Social media**
- Facebook 66%

**Online retail**
- Amazon 37%

Source: Global Stats Counter
Economist.com
It's Not Too Late for Social Media to Regulate Itself
The EU General Data Protection Regulation (GDPR) is the most important change in data privacy regulation in 20 years.

The regulation will fundamentally reshape the way in which data is handled across every sector, from healthcare to banking and beyond.
### CLEAR LANGUAGE

<table>
<thead>
<tr>
<th>TODAY</th>
<th>TOMORROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Often businesses explain their privacy policies in lengthy and complicated terms</td>
<td>Privacy policies will have to be written in a <strong>clear, straightforward language</strong></td>
</tr>
</tbody>
</table>

### CONSENT FROM USER

<table>
<thead>
<tr>
<th>TODAY</th>
<th>TOMORROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Businesses sometimes assume that the user's silence means consent to data processing, or they hide a request for consent in long, legalistic, terms and conditions — that nobody reads</td>
<td>The user will need to give an <strong>affirmative consent</strong> before his/her data can be used by a business. Silence is no consent</td>
</tr>
</tbody>
</table>
### More Transparency

<table>
<thead>
<tr>
<th>TODAY</th>
<th>TOMORROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>The user might not be informed when his/her data is transferred outside the EU</td>
<td>Businesses will need to <strong>clearly inform</strong> the user <strong>about</strong> such <strong>transfers</strong></td>
</tr>
<tr>
<td>Sometimes businesses collect and process personal data for different purposes than for the reason initially announced without informing the user about it</td>
<td>Businesses will be able to collect and process data only for a <strong>well-defined purpose</strong>. They will have to inform the user about new purposes for processing</td>
</tr>
<tr>
<td>Businesses use algorithms to make decisions about the user based on his/her personal data (e.g. when applying for a loan); the user is often unaware about this</td>
<td>Businesses will have to <strong>inform</strong> the user <strong>whether the decision is automated</strong> and give him/her a possibility to contest it</td>
</tr>
<tr>
<td>TODAY</td>
<td>TOMORROW</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Often businesses do not inform users when there is a data breach, for instance when the data is stolen</td>
<td>Businesses will have to <strong>inform</strong> users without delay in case of harmful data breach</td>
</tr>
<tr>
<td>Often the user cannot take his/her data from a business and move it to another competing service</td>
<td>The user will be able to <strong>move</strong> his/her <strong>data</strong>, for instance to another social media platform</td>
</tr>
<tr>
<td>It can be difficult for the user to get a copy of the data businesses keep about him/her</td>
<td>The user will have the right to <strong>access</strong> and get a copy of his/her data, a business has on him/her</td>
</tr>
<tr>
<td>It may be difficult for a user to have his/her data deleted</td>
<td>Users will have a clearly defined “<strong>right to be forgotten</strong>” (right to erasure), with clear safeguards</td>
</tr>
<tr>
<td>TODAY</td>
<td>TOMORROW</td>
</tr>
<tr>
<td>---------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Data protection authorities have limited means and powers to cooperate</td>
<td>The <strong>European Data Protection Board</strong> grouping all 28 data protection authorities, will have the powers to provide <strong>guidance</strong> and <strong>interpretation</strong> and adopt <strong>binding decisions</strong> in case several EU countries are concerned by the same case</td>
</tr>
<tr>
<td>Authorities have no or limited fines at their disposal in case a business violates the rules</td>
<td>The 28 data protection authorities will have harmonised powers and will be able to <strong>impose fines</strong> to businesses up to 20 million EUR or 4% of a company’s worldwide turnover</td>
</tr>
</tbody>
</table>
CALIFORNIA CONSUMER PRIVACY ACT OF 2018
Building on “FIPP”

Treat companies as “information fiduciaries” with basic duty of care

Designate platforms as “essential services” and ensure they provide services on “fair, reasonable, and non-discriminatory terms”

Some revamped model of public utility regulation

Create a “public trustee” – new agency to place obligations or duties on organizations (rather than individuals)

insulate from industry; avoid politicization

ex. models:

data protection agencies (Europe)
privacy commission (Canada)
Alternative: Civil Society Activism

Electronic Frontier Foundation

American Civil Liberties Union

Worker activism inside companies

=> Become a “public interest technologist” (Create the next Signal!)
Establish and enshrine a new right to privacy.

Try adding it to the Bill of Rights. “Congress shall make no law...”

What is the right?
To whom does it apply?
What is prohibited?
Are there any exceptions?
The Time For Debate is Now

The prospects for bipartisan cooperation on comprehensive privacy legislation are reasonably good.

A super majority of registered voters want a bill in this Congress.

Last Congress, there were bills on both sides of the aisle.

Ro Khanna says it is a top priority for the Dems who are in charge.

But neither Jennifer nor Steve sees legislation as likely in the next four years!
Today’s Agenda

1. Technical considerations in data privacy
   • Data storage
   • Multi-party privacy
   • Information discovery
   • Issues in facial recognition
2. Regulatory considerations in data privacy
3. Philosophical considerations in data privacy
Privacy is a value that is necessary to protect our higher-order interest in autonomy and self-determination.

*The claim of individuals to determine for themselves when, how, and to what extent information about them is shared with or communicated to others.*

**QUESTION:** How should we institutionalize – or put into practice – this value?
Back to Square One: What is Privacy?

Should we construe privacy as:

- An **interest**? (one consideration among many; can be defeated by other values or rights?)

- A **right**? (strong presumption against infringement)

- A **constitutional right**? (ordinary lawmaking can’t remove; sits above majoritarian decisionmaking)

- An **inalienable constitutional right**? (ordinary lawmaking can’t remove and right holder can’t renounce – no notice and consent!)

- A **fundamental human right**? (no government can remove/abridge)
Rival Values

What other values (or interests, or rights) might be in tension with privacy (however we construe it)?

1. National Security (e.g., terrorism)
2. Public Safety (e.g., crime)
3. Innovation
4. Convenience
Privacy vs. Innovation?

How GDPR Kills The Innovation Economy

By John Battelle

Dept. Of Predictable Externalities

It wasn’t supposed to turn out this way.
Surveillance Capitalism

"An intensively researched, engagingly written chronicle of surveillance capitalism's origins and its deleterious prospects for our society."
—JACOB SILVERMAN, New York Times Book Review

"The rare volume that puts a name on a problem just as it becomes critical."
—FRANK ROSE, Wall Street Journal

"An original and often brilliant work, and it arrives at a crucial moment, when the public and its elected representatives are at last grappling with the extraordinary power of digital media and the companies that control it."
—NICHOLAS CARR, Los Angeles Review of Books

"A masterpiece of rare conceptual daring, beautifully written and deeply urgent."
—ROBERT B. REICH, author of The Common Good and Saving Capitalism: For the Many, Not the Few
Opinion

You Are Now Remotely Controlled

ance capitalists control the science and the scientists, the secrets and the truth.

By Shoshana Zuboff
Ms. Zuboff is the author of “The Age of Surveillance Capitalism.”
We Need a Law to Save Us From Dystopia

It’s not too late. And it better be comprehensive.

By Charlie Warzel
Mr. Warzel is an Opinion writer at large.

Jan. 21, 2020
Opinion | THE PRIVACY PROJECT

Be Paranoid About Privacy

We need to take back our privacy from tech companies — even if that means sacrificing convenience.

By Kara Swisher
Ms. Swisher covers technology and is a contributing opinion writer.

Dec. 24, 2019
EVERY MINUTE OF EVERY DAY, everywhere on the planet, dozens of companies — largely unregulated, little scrutinized — are logging the movements of tens of millions of people with mobile phones and storing the information in gigantic data files. The Times Privacy Project obtained one such file, by far the largest and most sensitive ever to be reviewed by journalists. It holds more than 50 billion location pings from the phones of more than 12 million Americans as they moved through several major cities, including Washington, New York, San Francisco and Los Angeles.

Each piece of information in this file represents the precise location of a single smartphone over a period of several months in 2016 and 2017. The data was provided to Times Opinion by sources who asked to remain anonymous because they were not authorized to share it and feared repercussions for doing so. The possessors of
How Your Phone Betrays Democracy

By Charlie Warzel and Stuart A. Thompson
DEC. 21, 2019
Total Surveillance Is Not What America Signed Up For

By The Editorial Board

Dec. 21, 2019
It is a federal crime to open a piece of junk mail that's addressed to someone else. Listening to someone else's phone call without a court order can also be a federal crime.

The Supreme Court has ruled that the location data served up by mobile phones is also covered by constitutional protections. The government can't request it without a warrant.

But the private sector doesn't need a warrant to get hold of your data. There's little to prevent companies from tracking the precise movements of hundreds of millions of Americans and selling copies of that dataset to anyone who can pay the price.
You Should Be Freaking Out About Privacy

Nothing to hide, nothing to fear? Think again.

Featuring Farhad Manjoo and Kara Swisher  Video by Adam Westbrook

Dec. 20, 2019
Rival Values

• How should we balance privacy against security/public safety/innovation/convenience?
• What weight, under what circumstances, should we attach to these interests?
• Who should decide the framework?
• Who should decide who decides?
“If men were angels, no government would be necessary. If angels were to govern men, neither external nor internal controls on government would be necessary. In framing a government which is to be administered by men over men, the great difficulty lies in this: you must first enable the government to control the governed; and in the next place oblige it to control itself.”

-- James Madison, Federalist Papers No. 51
Tuning the System vs. Attacking the System

Should we be satisfied with tweaking the system to make it fairer or to protect privacy? Should we “tune” the system to produce different outcomes? Why not tackle the rigged system instead?

Where People Think the Economy is Rigged
Citizens of selected countries that think the economy is rigged in favor of the rich & powerful

- Mexico: most (94%)
- South Korea: 83%
- Russia: 80%
- United States: 78%
- Germany: 77%
- Great Britain: 76%
- Australia: 75%
- Canada: 71%
- Japan: 71%
- Sweden: least (56%)

Survey of 17,180 adults in 22 countries in September & October 2016
Source: Ipsos
During the last two decades, the leading surveillance capitalists — Google, later followed by Facebook, Amazon and Microsoft — helped to drive this societal transformation while simultaneously ensuring their ascendance to the pinnacle of the epistemic hierarchy. They operated in the shadows to amass huge knowledge monopolies by taking without asking, a maneuver that every child recognizes as theft. Surveillance capitalism begins by unilaterally staking a claim to private human experience as free raw material for translation into behavioral data. Our lives are rendered as data flows.

Shoshana Zuboff’s *The Age of Surveillance Capitalism*