

Taking Control of Your Personal Data

Course Guidebook

Professor Jennifer Golbeck
University of Maryland, College Park





4840 Westfields Boulevard | Suite 500 | Chantilly, Virginia | 20151-2299

[PHONE] 1.800.832.2412 | [FAX] 703.378.3819 | [WEB] www.thegreatcourses.com

LEADERSHIP

Paul Suijk	President & CEO
Bruce G Willis	Chief Financial Officer
Joseph Peckl	Senior Vice President of Marketing
Cale Pritchett	Vice President of Marketing
Jason Smigel	Vice President of Product Development
Debra Storms	Vice President, General Counsel
Mark Leonard	Vice President of Technology Services
Kevin Manzel	Senior Director of Content Development
Gail Gleeson	Director of Business Operations & Planning
Kevin Barnhill	Director of Creative

PRODUCTION TEAM

Trish Golden	Producer
Michelle Pellatt	Content Developer
Juliet Riley	Associate Producer
Lisa Persinger Robertson	Associate Producer
Trisa Barnhill	Graphic Artist
Daniel Rodriguez	Graphic Artist
Owen Young	Managing Editor
Art Jaruphaiboon	Editor
Charles Graham	Assistant Editor
William DePaula	Audio Engineer
Gordon Hall IV	Audio Engineer
Valerie Welch	Production Assistant
Roberto de Moraes	Director

PUBLICATIONS TEAM

Farhad Hossain	Publications Manager
Blakely Swain	Copyeditor
Tim Olabi	Graphic Designer
Jessica Mullins	Proofreader
Erika Roberts	Publications Assistant
Renee Treacy	Fact-Checker
William Domanski	Transcript Editor & Fact-Checker

Copyright © The Teaching Company, 2020

Printed in the United States of America

This book is in copyright. All rights reserved. Without limiting the rights under copyright reserved above, no part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording, or otherwise), without the prior written permission of The Teaching Company.

Jennifer Golbeck, PhD

**Professor in the College
of Information Studies**

**University of Maryland,
College Park**



Jennifer Golbeck is a Professor in the College of Information Studies and Director of the Social Intelligence Lab at the University of Maryland, College Park. She received an AB in Economics and an SB and SM in Computer Science at the University of Chicago, as well as a PhD in Computer Science from the University of Maryland, College Park.

Professor Golbeck began studying social media from the moment it emerged on the web, and she is one of the world's foremost experts in the field. Her research has influenced industry, government, and the military. She is a pioneer in the field of social data analytics and discovering people's hidden attributes from their online behavior, and she is a leader in creating human-friendly security and privacy systems.

Professor Golbeck is the author of several print and online publications, including the book *Analyzing the Social Web*, and she is a frequent contributor on NPR. Her TED talk on what social media likes can reveal about you was featured in TED's 2014 Year in Ideas. ■

TABLE OF CONTENTS

INTRODUCTION

Professor Biography	i
Course Scope	1

LECTURE GUIDES

Lecture 1 How Your Data Tells Secrets	2
Lecture 2 The Mechanics of Data Harvesting	11
Lecture 3 Privacy Preferences: It's All about You	22
Lecture 4 The Upside of Personal Data Use	32
Lecture 5 Online Tracking: Yes, You're Being Followed	42
Lecture 6 Nowhere to Hide? Privacy under Surveillance	53
Lecture 7 Consent: The Heart of Privacy Control	63
Lecture 8 Data Scandals and the Lessons They Teach	73
Lecture 9 The Dark Web: Where Privacy Rules	83
Lecture 10 Algorithmic Bias: When AI Gets It Wrong	92
Lecture 11 Privacy on the Global Stage	101
Lecture 12 Navigating the Future of Personal Data	111

SUPPLEMENTARY MATERIALS

Resources	122
Image Credits	128

TAKING CONTROL OF YOUR PERSONAL DATA

Today's headlines are dominated by reports of scandals and hacks involving millions of people's personal data. There has never been a more important time to be informed and in command of what personal data is, how it's used, and what you can do to protect your privacy.

This course introduces you to the complicated and evolving world of personal privacy in the digital age. You'll learn what constitutes personal data, both online and in your physical interactions. You're already familiar with much of this data and how it's gathered and used, but you may be surprised at how data can be aggregated and analyzed by powerful algorithms to reveal astonishing things about you.

You'll be guided through the mechanics of the many technologies that exist to track your behavior online. Sometimes these technologies make your life easier, but often they are used in ways that are hidden and can be used to manipulate you. How comfortable you are with these practices is a personal choice. This course will help you figure out where you stand on the spectrum of privacy approaches so that you can choose the strategy that works for you.

The course offers concrete advice and tools for creating and maintaining privacy controls that work for you. You'll find a variety of control strategies for all of your privacy concerns, from VPNs and alternative social media options to technologies that help you navigate the bewildering world of consent. Some of the most serious personal data issues can only be addressed through regulation, and you'll discover how governments in the US and abroad are dealing with these serious concerns.

The world of personal data is constantly shifting as technology evolves. This course teaches you how to ask the right questions and find the right answers to keep your data safe and protected—today and in the future.



HOW YOUR DATA TELLS SECRETS

LECTURE 1

The landscape of data collection and analysis has changed many times over the years—sometimes in expected ways and sometimes in surprising ways. Never before in human history have we been able to share so much about ourselves with other people so quickly. And never in our history have we been so exposed to forces that want to take advantage of that ability.

And this new reality has pros and cons.



Artificial Intelligence Algorithms

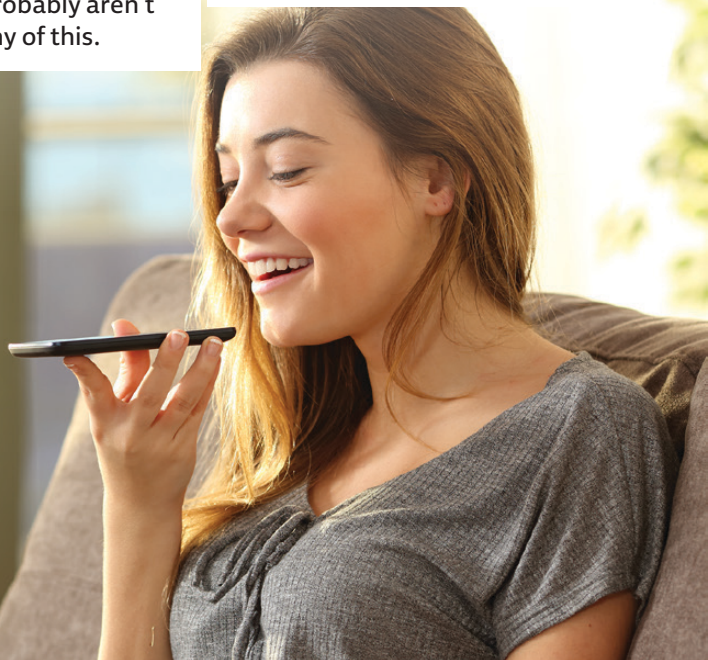
- Personal data can be a powerful tool to help people. Consider this example. Suicides among soldiers, especially those returning from conflict zones, is unacceptably high. The Department of Defense acknowledges this as one of their most frustrating problems that they want to do more to address. One project showed how soldiers' social media data and other writings could be analyzed by artificial intelligence and used to identify people at risk of self-harm. The resulting tools could identify suicidality more accurately than state-of-the-art medical approaches, and these were built into other tools to help medical professionals intervene.
- But that same analysis can also be used to hurt people. One infamous story comes from Canada, where a woman was diagnosed with severe depression and put on long-term sick leave. Her doctors advised her to try to have fun and maybe take a vacation.

When she did, she posted a photo of herself at the beach on Facebook. The insurance company paying for her disability saw the photos, decided that she looked like she was having a good time and thus was no longer depressed, and revoked her coverage. Apparently no one told the insurer that even severely depressed people can smile and laugh sometimes.

- We all know that when we post things online—whether it's a restaurant review or a social media post—this personal information can be seen and shared. Data is collected in many startling ways, and new information is created from that with artificial intelligence.

Our data tells stories. It is bad at keeping secrets. Artificial intelligence can bring those secrets out—sometimes for good and sometimes in very creepy ways.

- ❑ You create a lot of data by just existing in a digital world. Your purchases are tracked through your use of credit cards and loyalty cards. Your internet traffic and web searches are recorded. Your social media posts contain all sorts of personal information. Your phone tracks your location, the apps you use, and the time of day you are using them and connects that with your contacts, calendars, phone calls, and text messages. Cameras capture your movements in public spaces. It can be overwhelming when you list it all out like that, but most people probably aren't surprised by any of this.
- ❑ But data is collected in many more subtle—and some might say nefarious—ways. Maybe you're chatting with a friend in person, not using your phone at all, and say, "Next year for spring break, we might go to Costa Rica!" You don't look up anything about Costa Rica or post about it; you only make the comment to your friend. Yet the next day, ads for Costa Rican tourism start appearing on Facebook or blogs or other websites. How does that happen? Some apps have code that turns on your phone's microphone and passively listens in!



- Huge amounts of data are being collected about you, and you're probably unaware of many of them. Why is so much collected? Because it can be used to generate even more information about you!
- You're probably familiar with artificial intelligence algorithms that use your data, even if you haven't thought about it. For example, Netflix and Amazon both make recommendations to users. Netflix suggests shows you might want to watch; Amazon recommends products you might want to buy. They do this with artificial intelligence algorithms that analyze what you have interacted with before and then guess about new things you might like in the future. But what's going on inside those algorithms, and what else can they find out about you?
- An algorithm is just a series of steps a computer takes. Artificial intelligence algorithms are steps a computer takes to become "intelligent" about something—such as Siri understanding what you're asking or a map application finding a good route for you to take.

Machine Learning Algorithms

- A machine learning algorithm is a specific kind of artificial intelligence algorithm that helps a computer learn things. As a simple example of how these algorithms work, say you have an app that tracks when you go outside for a walk or a run. It also knows the weather. If that app had a machine learning algorithm built in, it could start to learn which weather conditions are conducive to your walks.

Each day, it would note the conditions and then wait to see if you go outside. If you do, it learns that that day's conditions are good. If you don't, it learns that they are bad. Over time, it builds up a model of your preferences.

There are great ways that algorithms help people every day, but there are also terrible uses of it that invade people's privacy and impact their lives.

- ❑ We build up mental models like this for ourselves and others all the time. If you walk with a friend, you probably know what days your friend will want to come out and what days he or she will want to stay home. That's because you have learned a model of his or her behavior, and even if you've never seen a day with this exact set of conditions, you can still make a good guess.
- ❑ Machine learning models do the same thing. The algorithm takes some input, such as weather conditions, and an answer about whether it is a good or a bad day. Then, the algorithm builds a model from a series of examples. When it's done, you can ask it a question about a totally new set of weather conditions, and it can make a guess about whether it's a good or a bad day.
- ❑ Machine learning models can use your personal data as input instead of weather conditions to build up models that let them guess all kinds of personal things about you.
- ❑ A lab at the University of Maryland did one of the first research projects in this space in 2012. Researchers used people's public Facebook information to guess what their personality traits were. Basic personality tests, such as the Myers-Briggs Type Indicator personality inventory, measure things like whether you're an introvert or extrovert, if you're a procrastinator or a planner, and if you're open to new experiences or prefer to do the same things habitually.
- ❑ To make an algorithm that could guess these traits from Facebook data, researchers had hundreds of people take a personality test and then fed a machine learning algorithm their Facebook data and their scores. The algorithm learned a model of each personality trait. In the end, the algorithm could guess someone's score on the personality test exceptionally well just by examining their Facebook data. It was only off by about 10 percent, which is basically how much your score can change when you take a test depending on the mood you're in.

□ Since that first study, there has been an explosion of work on this topic. In a study conducted at Cambridge University, researchers wanted to see if they could use Facebook likes to predict a whole host of personal attributes.*

They used the likes to predict demographic information, such as race, religion, gender, and sexual orientation; behavioral attributes, such as if you were a drinker, smoker, or drug user; and preferences, such as if you were liberal or conservative. They could even tell if your parents divorced before you were 21.

□ You may be thinking that you don't need any algorithms to connect likes to some of these attributes. If you like the GOP page on Facebook, there's a good chance you're conservative. If you like the Jim Beam page, there's a good chance you're a drinker.

But it turns out that algorithms don't rely on these obvious indicators very often. Instead, they are finding statistical connections that don't make much sense to humans.

□ These algorithms are very opaque. They can give us an accurate answer a lot of the time, but they don't give us any insight about how they got to that answer. Part of that is intentional. Computer science as a field is only concerned with whether something can be computed; it is not interested in the social insights as to why.

Research has found that viral videos, fads, rumors, fake news, and pages to like on Facebook all spread in online social networks the same way that diseases spread in offline social networks.

* On Facebook, you can like all sorts of things—sports teams, movies, books, bands, restaurants, or any page that a person creates. The things a person likes is a pretty narrow sliver of all the information that person shares on Facebook, but it is always public. You can't make your likes private, so anyone can get access to them at any time. In other words, people can use your public likes to discover private information about you.

Homophily and Correlations

- ❑ Most of what you like on social media is social—it's stories, pages, pictures, and videos that appear in front of you because someone shared it. Rarely is it something you go search for yourself. So what you see depends on your friends.
- ❑ Sociologists have known for a long time that we are friends with people who are similar to ourselves. They call this

homophily, and it's basically the concept that birds of a feather flock together. If you're rich, your friends tend to be rich. If you graduated from college, your friends also probably graduated from college. It's not that all of your friends are the same as you, but rather that your personal attributes are shared by your friends more often than they are shared by the general population.



- Essentially, an algorithm guesses that you belong in the group of people who liked a certain page, and that group's traits are reflected back onto you. Without knowing it, the algorithm is picking up on traces of your social interactions and using them to learn more about you. Homophily is incredibly powerful when used by these algorithms. It's important to know that these are not meaningful correlations in the data from a human perspective. They are statistical patterns that are found for some people in the context of millions of others that allow a mathematical tool to separate people into groups.
- It's also important to realize that you can't hide from algorithms. You may be able to control what you post on social media or what you buy, but you don't know what your actions say to an algorithm. So if there is something you want to keep private, there's no way to know how to do that when an algorithm is doing the analysis.
- Also, correlations found by algorithms are not stable. Algorithms constantly refine and rebuild their models and don't rely on specific correlations explicitly. For users, it means that you can't adjust your actions based on knowledge about what was predictive last week, because it may have changed.
- In addition to algorithms discovering things that are true about you right now, researchers are getting very good at predicting things that will be true about you in the future—even before you know it.
- One project followed pregnant women on Twitter and developed a model that could predict whether they would develop postpartum depression. On the day a woman gives birth, the algorithm already knows the outcome—and it is right 80 to 85 percent of the time. It does this using data about the style of people's writing, their interactions, and the kinds of posts they make.

So What Can You Do?

- ❑ How can you take control over these algorithms? In most ways, you can't.
- ❑ In the US, you do not own your personal data, and companies do not have to tell you what they know about you or how they are using your data. These algorithms are used behind the scenes all the time, often without your knowledge.
- ❑ Sometimes it is for relatively benign purposes, such as showing you the most relevant ad. Sometimes it's really helpful, such as Amazon notifying you of items you need to go with a new appliance you bought. But they were also used in the 2016 presidential election to try to suppress and manipulate the vote—and no one told you about it until after the fact.

Assignment

Where are you leaving data behind? Any time you participate in a digital transaction—log onto a website, buy something with a credit card, drive through a toll sensor, or check the weather on your phone—data about you is collected. Over the course of these lectures, you will probably discover new ways you are being tracked that you didn't know about! But for now, try to make a comprehensive list. Don't forget to include any device in your house that has internet access, such as a scale, smart light bulb, or video doorbell. You'll probably be surprised by how long your list is.

Resources

Duhigg, "How Companies Learn Your Secrets," <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

Golbeck, "Smart People Prefer Curly Fries," <https://slate.com/technology/2014/10/youarewhatyoulike-find-out-what-algorithms-can-tell-about-you-based-on-your-facebook-account.html>.

Netflix Research, "Recommendations," <https://research.netflix.com/research-area/recommendations>.



THE MECHANICS OF DATA HARVESTING

LECTURE 2

When thinking about being careful with our personal data, especially online, we tend to think about what we explicitly choose to share, such as our address, phone number, or other sensitive personal information. But data is collected about us in many other ways. Some of these we know about. Our phone companies keep track of who we text on our mobile devices; our credit card companies keep track of where we shop and what we spend. In addition, though, vast amounts of data are being collected that you might not know about. Websites and your devices with apps you've installed are regularly collecting huge amounts of information about you, including sensitive information that they probably don't need to operate.

How—and How Much—Information Is Collected

- Facebook wants to know details of how people use their platform. Their main goal is to keep people engaged with it. They want people to use Facebook as much as possible and to know what kinds of activities keep people engaged. For example, if you are commenting on your friends' posts, that is good from Facebook's perspective because it keeps you on Facebook and encourages your friends to engage. But what if you start to post something and then reconsider, never actually posting it?
- Facebook wants to know when this happens, so they have code that collects information about you when you do this. We don't know if they are collecting the text you type or just data like how much you typed, but the technology exists that would let them grab your comment, store it, and analyze it—even if you don't actively post it.
- When this was first reported, Facebook claimed that they were not collecting what people typed in the box, just how much they typed there. But if their policies changed, or if other services used this technology, they could store that information and use it in a variety of ways. The technology to collect text like this is very straightforward to use. Thus, it is safe to assume that many websites are harvesting information that you type on their platforms, even if you choose not to send it.
- Your phone is also a gold mine for companies who want to collect information about you. For example, Geoffrey Fowler at *The Washington Post* wondered about his iPhone. Just how much data was it sending out that he didn't know about? He worked with a company called Disconnect to monitor the data that was sent from his phone and was shocked by the results: In a week, 5,400 hidden apps and trackers received data from his phone.

During the Women's World Cup, La Liga, a National Professional Football League in Spain, used their app to spy on users. They turned on the microphone to listen in and hear if the user was in a bar or other establishment that had the game on TV. Then, using location services, they could identify exactly which establishment that was. If the bar didn't have a license to show the game, La Liga could come after them. Essentially, users' phones were hijacked to catch bars showing the game who hadn't paid for the rights. They were able to do this because users allowed microphone and location access to the app when they installed it, but the app did not say what it planned to do with that access.



- Apps, trackers, and your phone itself can track your location, your pattern of movements, who you call and how long you talk, who you text with and how often, what other apps you have installed, your phone number, your contacts, and sometimes even your photos. And in most cases, we do not know who these companies are or what they are doing with our data.
- Some of them are tracking us in especially surprising ways. For example, some apps turn on your phone's microphone and passively listen in on what's happening in the background. They can use this to pull out keywords you say or to identify TV shows or songs or commercials you are hearing. All of this feeds into a profile about you, your interests, and your activities.



□ Even without access to the microphone, though, it is possible to listen in on you. Phones have accelerometers that can tell how fast you are moving in three dimensions. This is used for things like the compass, fitness apps that count steps, and games you control by tilting your phone. There is no special privacy permission that controls access to the accelerometer. Researchers have shown that the sound from talking causes vibrations that the accelerometer picks up, and an app could analyze those and convert it into speech.

□ And even when we do know that our devices are listening, they may capture more than we expect. Virtual personal assistants work by listening for a “wake” word or phrase, such as “Hey Siri” or “Hey Google” or “Alexa.” They record what follows, upload that audio to the host (such as Apple or Amazon or Google), where it is processed and analyzed, and a response is generated and sent back. If someone were to activate an Echo while something criminal* was happening, Amazon may well have a recording of it.

* Timothy Verrill is accused of murdering Christine Sullivan and Jenna Marie Pellegrini in January 2017. On the night of the murder, he broke into Sullivan’s home and brutally beat and stabbed the two women to death. There was an Amazon Echo personal assistant device in the home, and prosecutors took the device. They believed there may be some recordings of the actual murder on it. A judge in the case also ruled that Amazon had to turn over any recordings they had.

❑ Perhaps you are not criminally inclined and therefore are not particularly worried about your personal assistant device incriminating you at trial, but the fact that these devices can and do make recordings in your home without your knowledge should set off alarm bells.

❑ Just how much does Amazon collect from devices like this? In late 2018, a German Amazon user requested an archive of all of his data that the company held, a right he has under the European privacy General Data Protection Regulation. Amazon sent 1,700 recordings from someone else—including some recordings that person had made while he was in the shower. The recipient contacted Amazon but never heard back from them. So eventually, he went to the magazine publisher Heise with what he had received.

❑ The journalists found that by listening to recordings that asked for weather and mentioned people's first and last names and friends' information, they were easily able to identify the voice on

the recordings as the man's girlfriend. Amazon says that releasing this data to the wrong person was a mistake, but the fact that they save thousands of recordings from people is an interesting fact on its own. These recordings contain deeply personal data that, when aggregated, can reveal a tremendous amount about a user.

❑ And why collect all that data? Because it can be used to profile you. The realm of things that can be understood from simple data is vast—and growing. In fact, recent work has been able to predict things that will be true about you in the future before you even know it.

A study conducted at Cornell University set out to identify someone's spouse or significant other on Facebook by looking only at which of their friends knew one another. As they analyzed their data, they accidentally discovered that they could predict whether someone's current relationship was likely to last or fall apart in the near term.

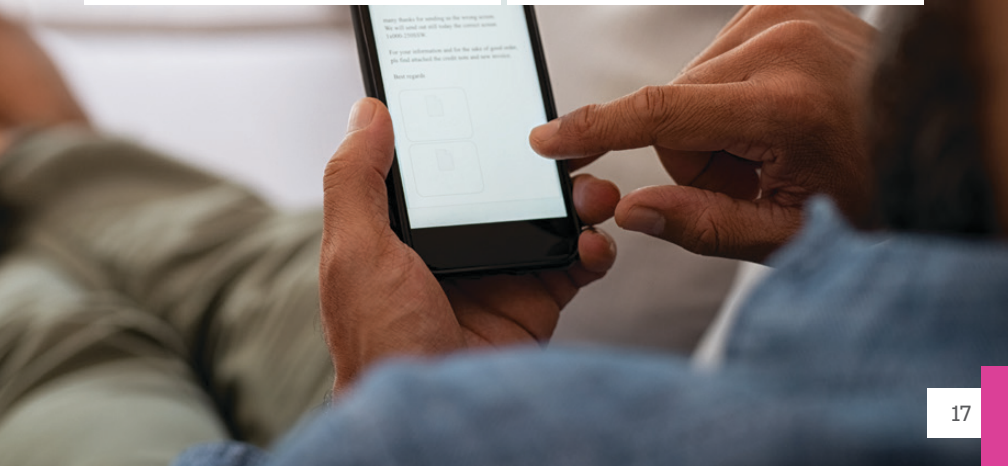
- ❑ Many researchers study how to build algorithms to predict future behavior and attributes. A lot of concern comes with this technology, especially when it operates on public data. It can really help people and the organizations they work with. At the same time, it can be incredibly intrusive and used in unfair ways. We don't yet know how to strike the balance.
- ❑ There's a lot of information that can be uncovered about you from background data that you may be unaware is being collected. On top of that, you're even less likely to be aware of the algorithms being applied to that data and the power they may have to understand things about you. And basically, you can't hide from them.

Shadow Profiling

- ❑ Even if you have not created a Facebook profile, Facebook still knows a lot about you. For many people who are not on the site, Facebook creates something called a shadow profile. Basically, it's very easy to know when a person is missing from a social network. There is a hole where a person should be, and Facebook can easily figure out who that person is that the hole represents.
- ❑ So if you don't have a Facebook account and you've never shared any information with Facebook, how do they build a profile about you, and what could they come up with?
- ❑ Facebook admits that they have these shadow profiles, but we don't know a lot about how much data they have in each one or how they calculate that data. But there is straightforward technology that could be used to build one of these shadow profiles. We don't know exactly how Facebook is doing it, but what follows is an educated guess. Also, the focus here is on Facebook, but most other big social media platforms could and may do this, too.

□ The most obvious information to use is other people's contact lists. If you have a friend who has a social media account and that friend uses the app, your friend likely has given access to his or her contact list. Many platforms ask for this because it allows them to pair you up with other people who are using the app. They do this by downloading a list of your contacts along with all of your contacts' data—their phone numbers, email addresses, street addresses, and maybe photos—and if a particular platform has another user in the system with that same email address or phone number, it can suggest that you become friends with that person. Essentially, a phone number or an email address are unique identifiers.

□ If you are not on Facebook but you are in someone's contact list when Facebook downloads your friend's contact list, Facebook now knows that you exist, what your name is, what your phone number is, what your email address is, and maybe things like your street address or your website or a picture of you. Getting that from one person is useful. But the vast majority of Americans who have internet access also have a Facebook account. So even if you are not on Facebook, most of your friends still are, and you are likely in many of their contact lists. And when your friends give permission for Facebook to access their contacts, Facebook retrieves your data from a lot of different people.



- ❑ So now Facebook doesn't just know that you exist. It also knows who a bunch of your friends are. And since these friends have profiles, some of them are likely friends with each other, while others are from different social circles. This can reveal your interests.
- ❑ For example, several people may be in your neighborhood's Facebook group. And if you have three or four people who you know in real life and those people have you in their contact list and are on Facebook and also part of the neighborhood group, Facebook may infer that you're likely to live in that same neighborhood, especially if Facebook has your street address and can show that you live nearby.
- ❑ Similarly, if you go to a church or temple or mosque and have friends on Facebook who have you in their contact list who go to that same religious institution and who are friends with each other, Facebook may be able to infer your religion from that. The same thing applies if a lot of your friends share an interest in a sports team or list the same employer. From this information that your friends have provided, Facebook knows your name, your location, your contact information, a bunch of your friends, and many of your interests.



□ This information can also reveal other traits. For example, research shows that it's quite easy to determine sexual orientation based solely on information about people's friends, especially for men. So even if you keep your sexual orientation private online, if you have

friends who do not, Facebook may be able to tell your sexual orientation just from information that other people provide. These details about people who have opted out of the system can be used in a variety of ways with both good and potentially dangerous consequences.

So What Can You Do?

□ Check out how much of your data is being shared. There are apps that will allow you to do this. The Privacy Pro app, created by the Disconnect team that helped *The Washington Post* investigation, has a free option that will let you monitor trackers and information being shared in the background on your phone. It will also block them, which speeds up the performance of your phone and protects data that would be shared.

□ Delete old posts on social media. When less information is available about you, algorithms can discover less about you as well.

Technical steps alone cannot stop the mass collection of data about us, but we can make it a lot harder to collect.

□ Check your privacy preferences. On your phone, turn off apps' permission to access your location, contacts, and other information, unless it is really critical. To stop background data collection, select the options that prevent apps from running in the background, and if they do not need to contact the internet—as is the case with many games—turn off their rights to use cellular data.

- By blocking apps from using data, you are preventing them from being able to send any information about you out to the world. This action may also stop them from downloading and showing ads to you. This might disable some features of the app, but you can decide which option feels best for you for each individual app.

Assignment

Check out your privacy settings. If you are on social media, such as Facebook, Instagram, or Twitter, log into your account, look at your privacy settings, and see what your default sharing is set as. Are you OK with it? Are you surprised by the setting? If you have a smartphone, go to your settings and select Privacy. Look at the different features and see what you share with which apps. Is this what you expected?

Are apps accessing more than you thought? If you don't have any social media accounts and don't have a smartphone, you can still look at your TV, cable box, or other household devices. Some of these will have privacy policies in the settings. You usually can't change these, but it is informative to see what your devices are collecting. This is a great opportunity for a privacy checkup—update any settings you don't like or that concern you and try the new settings for a while. Your privacy may improve without any negative impacts.

Resources

Fowler, "It's the Middle of the Night. Do You Know Who Your iPhone Is Talking To?" <https://www.washingtonpost.com/technology/2019/05/28/its-middle-night-do-you-know-who-your-iphone-is-talking/>.

Haskins, "Amazon Sent 1,700 Alexa Recordings to the Wrong Person," https://www.vice.com/en_us/article/pa54g8/amazon-sent-1700-alexa-recordings-to-the-wrong-person.

Matias, "Spain's Top Soccer League Fined for Using App to Spy on Fans in Fight to Curb Piracy," <https://www.npr.org/2019/06/12/732157537/spains-soccer-league-fined-for-using-app-to-spy-on-fans-in-fight-to-curb-piracy>.

Perez, "Some Apps Were Listening to You through the Smartphone's Mic to Track Your TV Viewing," <https://techcrunch.com/2018/01/02/some-apps-were-listening-to-you-through-the-smartphones-mic-says-report/>.

Whittaker, "Judge Orders Amazon to Turn Over Echo Recordings in Double Murder Case," <https://techcrunch.com/2018/11/14/amazon-echo-recordings-judge-murder-case/>.



PRIVACY PREFERENCES: IT'S ALL ABOUT YOU

LECTURE 3

When it comes to personal data and our individual privacy concerns, there is no one-size-fits-all solution. Every one of us falls somewhere along a spectrum regarding how much information we are willing to share and in exchange for what. Part of taking control of your personal data involves getting to know where you stand on that spectrum.




The Privacy Spectrum

- One easy and useful way to measure where you stand on the privacy spectrum is with the Westin-Harris privacy scale,^{*} which groups people as privacy fundamentalists, privacy pragmatists, and privacy-unconcerned individuals. The scale has three statements and asks people to rate them on a scale from 1 (strongly disagree) to 5 (strongly agree):
 1. People have lost all control over how personal information is collected and used by companies.
 2. Most businesses handle the personal information they collect about users in a proper and confidential way.
 3. Existing laws and organizational practices provide a reasonable level of protection for user privacy today.
 - On question 1, a score of 4 or 5 indicates concern, while a score of 1 or 2 indicates a lack of concern. On questions 2 and 3, the reverse is true, with a score of 1 or 2 indicating concern and 4 or 5 indicating a lack of concern.
 - Privacy fundamentalists want to closely control their data; they give answers that indicate concern on all three questions. Privacy-unconcerned individuals generally do not worry about their data and give answers that indicate a lack of concern on all three. Privacy pragmatists consider the trade-off between privacy and convenience, and their scores fall in the middle.
- About 10 percent of people are unconcerned, 25 percent are fundamentalists, and the rest are pragmatists.

^{*} The scale was originally developed to measure people's concerns about financial information, such as credit card numbers getting out, but has been modified to help understand concern about personal data more broadly.

- ❑ **Knowing where you fall on the scale can guide your decisions.** If you are a privacy fundamentalist who doesn't want any of your data collected, used, or analyzed without your explicit permission, you are likely to want to take more steps to protect it.
- ❑ **Privacy pragmatists** will have to think about how much value they get out of applications that collect their data and where the trade-off line is for them between the convenience and features they get and protecting their privacy.



As a society, we have by default made social media a repository for our lives, but there is no reason that you need to have years of personal information stored on Facebook or Twitter or Instagram.

Fortunately, there are automatic tools that can help you purge your social media content or periodically review it with targeted deletion.

- Facebook Timeline Cleaner runs in your browser, and you can specify that it should delete everything older than a certain number of hours.
- Tweet Delete automatically deletes content that's more than a specified number of days old.

If you're interested in adopting this strategy but don't want to lose all the content you have created, all the major social media sites have an option in the settings that allows you to download an archive of your data, including all of your posts, photos, and comments.

- Being unconcerned about your data is OK if you truly understand all the ways that your data can be used and you genuinely don't worry about it.
- It's also useful to think about who you are worried about having your data and using it, because that can also guide you to think about where you want to put your efforts into controlling your data. To help you understand this, imagine three people:
 - 📖 A person you're going on a date with. This is someone who's probably going to search for you on Google and Facebook and may spend 10 minutes looking through the information that's out there online. This person is not trying to stalk you or get some deep understanding of your psyche; he or she just wants to understand a little bit more about you.
 - 📖 A lawyer representing your ex in a vicious custody battle. This person is willing to spend a lot of time and money to dig very deeply into the information that is available about you. He or she may even employ several people to spend many hours combing the internet for any little crumb that may be able to be used against you.
 - 📖 A CIA agent. This person has access to tremendous power, resources that are not available to the general public, and technology to do things that even sophisticated people or organizations cannot. And this person is likely pursuing an agenda that you are not aware of.
- These three imaginary people represent three broad types of people or organizations that might gather and use your data.
- The first group, represented by your potential date, is relatively innocuous—not searching deeply, no harmful intent, just seeking to get to know you a bit using data that is readily available. This group might also include a potential landlord or employer.

- The second group, represented by the lawyer, may be more worrisome. This person has access to the time and resources needed to dig deeper and is motivated to act against you. Scammers and small-scale hackers might also be in this group.
- The third group, represented by the CIA agent, may be most troubling of all, with vast resources available and whose presence and motives may be completely invisible to you. This group would include government organizations who are doing an investigation, but also large technology companies who are willing to collect, analyze, and store data about hundreds of millions of people to make money off them. This group doesn't necessarily have to have negative intent toward you, but it's a group that has much more power than an average person to collect information about you and to analyze it.
- For each type of data, you have to think about which of these groups you're worried about. It may be all. But if you're not worried about an organization as powerful as the government, then you may be able to skip some steps that will protect your data. Often, the more steps you take to protect your data, the more inconvenience you have to deal with. This may come from slower connection speeds, loss of functionality, or extra steps you have to take in the course of your work. Your personal preferences will dictate what you want to put up with and when you're willing to give away information in exchange for convenience.

Types of Data to Control

- There are three types of data that you want to be able to control: data that you are explicitly choosing to share, data that is collected about you in the background that you may not know about, and data that is inferred or calculated about you using artificial intelligence.

We can be lulled into thinking that even if our posts are public and available to everyone, the only people who are really looking at them is a small group of friends. But studies have shown that people tend to vastly underestimate the number of people who actually look at their posts.

□ The easiest group to tackle is data that you explicitly share. This tends to be where you get the most guidance about taking care on social media: Be careful about how much you post, don't share your address, and maybe turn off location ID.


□ All that traditional advice is good if you want to keep this information private. And again, it's a reasonable and sane position if you say that you don't care. The one caveat is just be careful when you're thinking about how you might be revealing other people's information. For example, consider who else is in the pictures you want to post and if they would be OK with them being posted.

□ If you want to control what can be discovered about you through what you explicitly share, here are some basic steps to take.

First, review the default privacy settings on social media and any other internet-based applications where you are storing your data, such as a photo-sharing or video-hosting website. The default setting for these is often to share everything publicly.

□ If you are using social media or other internet-based services with the expectation that only your friends are looking at your posts, then you should make sure your privacy settings actually limit the visibility of your content to those people. If you don't limit it, it is likely that many more people are seeing it and thinking about how to use it. On pretty much every platform, there will be a section for privacy and security under settings where you can limit the visibility of your content.

- Some sites have very sophisticated privacy controls, but the basic choice that you will have is to make everything public—which is often the default—or to limit the visibility of your posts, which usually means only friends or people that you approve can see your content.
- What about information that is collected about you in the background? This can be data that is harvested off your phone or collected from your web-use behavior. This gets much more complicated and is a good place to spend time thinking about where you fit on the privacy spectrum and how you want to control your data. For information that's being collected in the background, we rarely know what that is or even who is collecting it and how it will be used.
- If you want to limit this kind of surreptitious data collection, what can you do? One tool that can be very powerful in combating background data collection is virtual private networks (VPNs). This technology prevents internet service providers (ISPs) from analyzing the websites you visit and aggregating and selling that information to advertisers—which is something that they are legally allowed to do. Basically, VPNs route all of your internet requests from your computer to a third party, which then sends them out to the internet.
- Another thing you'll have to think about in establishing your privacy profile is how you use other connected devices. We pretty much all have internet access on our computers and phones, but there are many other devices in our lives that want and may use internet access. And if they have internet access, it's likely that they are collecting information about you. Just how extensive that information is depends on the device.
- In-home surveillance devices, such as an Amazon Echo, can collect extensive recordings of your personal conversations, along with a lot of other data. Smart devices,



such as smart refrigerators and TVs, often come with privacy warnings to not have sensitive conversations in front of the devices because they could listen in and these conversations could be captured.

- Once again, however, there is convenience that comes with having these internet-enabled devices. What kind of privacy you are willing to trade for that convenience is up to you. But you should make a conscious decision about this from a privacy perspective instead of just allowing this technology in that may be eavesdropping or spying on you.
- The final important step in figuring out where you stand on the privacy spectrum is knowing how information gets inferred about you from your personal data—even if you are careful about what you choose to share on social media and you have made informed choices about what you share.
- Plenty of insights about you and your behavior can be inferred from even seemingly anemic data about you. On the one hand, you don't have a lot of control over this. Once the data about you is out there and people get a hold of it, they can basically do whatever they want with it, including making all sorts of inferences about you.

- The one place you do have power is in limiting what they are able to use. The most powerful groups, such as governments and very large tech companies, are likely to amass this data about you no matter what. There's very little you can do

to hide it from them beyond the steps that already have been discussed. But you do have an element of control in choosing the platforms on which you share information. Some are more privacy-conscious than others, so shop around.

So What Can You Do?

- Treat social media as a more ephemeral store of information, and use alternatives where it makes sense. For example, for sharing photos, you might currently use Facebook, which makes it easy to share those photos with friends and family. However, Facebook makes money by analyzing and exploiting your data, and there have been many instances where they have not been good stewards of that data.

- If you want to limit what Facebook can analyze about you and the people you know, you might want to look for an alternative photo-hosting service. There are many of these online that don't make any money off analyzing your photos or identifying people in them or learning anything from them. They may

charge you a small monthly fee to host the photos, but it's a small price to pay to have a private, password-protected site for hosting your photos.

- You can actually use a photo-sharing site in combination with Facebook. For example, instead of posting photos directly to Facebook, you could post a message to Facebook that says "the birthday party photos are now available on this website" with a link to a password-protected photo-sharing site. Facebook is unable to access or analyze those photos, and unauthorized people are unable to see the photos, but all of your family and friends on Facebook to whom you have given permission are able to log in and see the pictures.

- Consider moving things like private messaging out of the direct message or private message service from a social media site into a text message or group chat. Text messaging is much more private and secure.

Assignment

If you have social media accounts, such as a Facebook account, go far back in your posts and see what's there. Are you happy with it? Does it make you feel a little weird? Would you like other people to read through that old content, or would you maybe like it gone? This is a good chance to check in on the data you are storing online. It's fine if you are happy with what's there, but if you're not, take this opportunity to delete some old content!

Resources

Facebook, "Accessing & Downloading Your Information," https://www.facebook.com/help/1701730696756992?helpref=hc_global_nav.

Golbeck, "How to Set Up a Virtual Private Network," <https://slate.com/technology/2017/02/how-to-set-up-a-virtual-private-network.html>.

TweetDelete, <https://tweetdelete.net/>.



THE UPSIDE OF PERSONAL DATA USE

LECTURE 4

Many of the ways our personal data is used and collected are pretty creepy. But the reason these technologies are still being developed by many people who are not evil conglomerates or profit-hungry corporations is because they can be used in ways that benefit you. And knowing those benefits can help you make sound choices in controlling your data.



Simple Applications of Personal Data

- In the early days of the internet, if you searched for a term, such as “pizza delivery,” the search results you would get would be the most popular page in the world that matched that term. For many searches, that’s good, but for pizza delivery, it probably isn’t.
 - If you live in a small town in Iowa, you’re not interested in what the world’s most popular pizza delivery website is; you’re probably interested in a pizza delivery website that delivers to your town in Iowa. This requires showing different search results to different people based on what we know about them.
 - Any website can tell roughly where you’re located by your IP address, the unique identifier that every computer gets when it connects to the internet. From there, the search engine can add location in as a factor to determine what search results you are shown.
- This is personalization based on your personal data, but it’s not all that creepy.
- This is very basic technology, and it doesn’t include anything especially personally identifiable about you. Moreover, you know why your location is being used: to help you find the most relevant information using basic data that is easily available.
 - This same kind of technology is also used to do things like show pages that are written in your language. If you are using a search engine in English, it is likely to show you more English pages, especially if you’re located in an English-speaking country. If you are in Russia and have set your language preference to Russian, you will probably see more Russian-language pages. Again, we basically understand how companies get this information about us, and it is not invasive or surprising that they have it.

- ❑ Information about you is used to personalize your web searches in more sophisticated ways sometimes. If you allow it, in many search engines, histories of things you've searched for in the past are used to help build a profile of your interests. That can, in turn, be used to adjust what results you may see for a search. This can be very helpful if you are searching for terms that might be ambiguous.
- ❑ For example, if you live in an area that is wild and has a lot of bear activity, if you search for the word "cubs" in the spring, you're probably looking for information about the animal. If you live in the city of Chicago, you're probably looking for information about the baseball team. Personalized searches will not restrict you to search results about only the animal or the baseball team, but they may prioritize results about one topic higher than others depending on your interest profile.
- ❑ Similarly, patterns in our behavior are used online to personalize our experience with social media. Social media sites may learn what kind of information we are most interested in by paying attention to what we click on, what we spend time reading, and what we share. For example, Pinterest, the social networking platform that allows you to save visual bookmarks and organize them, uses patterns of items that you have interacted with in the past to show you new content.
- ❑ The more information sites use to personalize what you see, the more it inches toward feeling invasive. However, in most of these examples, the creepiness factor is relatively low. This is because we tend to know how and why the personalization is happening.
- ❑ We know how Pinterest, or any other social media platform, gets data about what we're interested in by tracking how we use their site. Most of us expect platforms to pay attention to what we click and save. And if they are only using that information to suggest things that we might want to see, we understand that this improves our experience and their service.

Dr. Munmun De Choudhury, a professor at Georgia Tech studying mental health and social media, develops artificial intelligence to understand people's emotional states from their social media posts with a particular emphasis on depression and suicide.

- Plus, even though your search engine has personal data about you, it is contained only within that company and used in a way that is straightforward, not surprising, and not particularly invasive. It helps you find more relevant information, and it also helps the company presumably make more money by offering a better service.

- Another important factor is that when a site uses information about their own users to personalize the content they see, users' personal information is not being shared outside of its use in personalizing their web pages. Your search engine might know where you are located and what language you are searching in, but it is not publicizing this to other people or companies.

Online Advertising

- The place we see personalization based on interests most often is in online advertising. Advertising is a gray area when it comes to personalization and whether it is good, bad, or outright evil.

There are some incredibly intrusive ways that advertisers monitor users' behavior; some of this surveillance crosses the line into activities that many people likely would consider illegal if they knew about them.

- ❑ However, if we accept that ads are a way that many of these websites support themselves and that there will be advertising one way or another, **some reasonable personalization, built on top of data that we know and expect a site to have, can not only make those ads better but even useful.**
- ❑ **Personalized ads come in many flavors. Some are personalized purely based on the content of the page.** If you are looking at a page about what kind of food to feed your dog, you are likely to get dog food ads appearing there. **They don't have anything to do with you.**
- ❑ **On web searches and on many web pages, the ads you see are almost always tailored based both on the content—the search term you used or the content of the page you are looking at—and on information about you, which may be something as simple as your location.**
- ❑ **If you search for farmer's markets, you will see search results that link to pages about** farmer's markets that are probably in your area, and the advertisements will tend to be for farmer's markets in your area as well. Both the content and the advertising is tailored to you. That could be very helpful if you are looking to find businesses.
- ❑ **The search results you get may not be a listing of farmer's markets but could be information about them or their history, whereas the ads will tend to be for actual markets where you can purchase things. If you're looking to buy stuff, it's helpful to get ads for companies who are selling things!**

In the early days of the web, advertising was placed similarly to how it is done in newspapers. If you went to a particular website, you saw an ad that an advertiser paid to have on that website to be seen by everyone who visited. It didn't matter who came; everyone saw the same ad. These types of ads still exist today, but they are much less common than the personalized ones.

Recommender Systems

- ❑ Companies who have a lot of personal data about us can also analyze it in deeper ways for mutual benefit beyond web-content personalization. This is something that your credit card company does. It knows a lot about you: where you are, where you often travel to, and how you normally spend your money. The company uses all of this information with algorithms to identify transactions that seem suspicious.
- ❑ Fraud alerts would not be useful if they were not personalized to each of our habits. They would either miss suspicious charges for some or send constant alerts to people who were going about their normal business. By using our data to personalize alerts, credit card companies can save us the hassle of dealing with fraudulent charges, and they can save themselves the money and effort that comes with resolving those cases. It's a win-win.



- Other examples of where this has been incredibly successful both for users and companies are in recommender systems: algorithms that suggest things to you that you might be interested in. You have probably encountered these on Netflix and Amazon. Netflix recommends shows and movies you might want to watch; Amazon recommends items you might want to buy. These types of systems generally do not feel creepy to people because we know how those websites got our data—by knowing what we interact with on their own platforms—and how they are using it—to help us find more of what they supply.
- We get useful information from these systems. There are far too many items for us to sort through ourselves. Netflix has thousands of shows and movies; Amazon has hundreds of millions of products. With that many items, it is helpful to have personalized suggestions about what we might want. When these systems work well, they bring consumers a lot of value by helping them identify items of interest.
- So how do these kinds of algorithms work with respect to your personal data? There are two major approaches: find things that are like the things that you like or find things that are liked by people who are like you. There are also hybrid approaches that combine these techniques.
- The first approach looks at things you like—movies, for example—and tries to find other movies that are similar. The only person whose information is needed is yours. The system relies on data they have about the items in their catalog to decide what to show you.

Netflix values their recommendation engine at 1 billion dollars; Amazon says that their recommendation engine is responsible for 35 percent of their profits.

- ❑ The other approach uses information from lots of people. This is called collaborative filtering. Such a system would look for other people who enjoyed the same movies you did and then recommend movies to you that were commonly liked among those people.
- ❑ Recommender systems also highlight an important aspect of how our data and the algorithms that use it should work. These are called decision support systems: systems designed to support you making a decision. They are not decision-making systems.
- ❑ For example, you probably don't just watch every show Netflix recommends. You don't let their recommender system decide what you should watch. It just supports you making your own decision about what to watch.
- ❑ Most of the time, these recommendation algorithms are correct in that they suggest things that you will probably like, but they are not especially insightful.
- ❑ Sometimes, the algorithm gets the recommendations totally wrong. It will suggest something that you would never enjoy. This tends to make up 10 to 20 percent of the recommendations on average. Most of the time, Amazon is recommending products that are basically the same as what you have already bought, but sometimes something weird will pop up into those suggestions.
- ❑ Finally, sometimes the algorithms get it right in a brilliant way that gives you a new kind of insight, or they help you discover something that you wouldn't have discovered before. If you have encountered this, it is probably through a streaming music service like Pandora or Spotify. You make yourself a channel and listen to songs that you expect, and then sometimes, it will play a song that you have never heard, but it's exactly what you like. Though this is a very small percentage of what the algorithms recommend, it is where they really show their value.

So What Can You Do?

□ How do you exert control over recommender systems? The answer depends on if you want to make them work better for you or if you want them to stop knowing so much about you and to be less effective.

□ If you want them to work better, providing more information is key. For recommender systems, that means rating more things, such as giving star ratings or liking things.

□ If you don't like this kind of personalization, you have a few options. You can rate fewer things and, where possible, delete the history of your browsing and searching.

You can also try to confuse the algorithm by giving it a lot of false information.*

□ If you want to use this approach, you can get some automated help. Noiszy is a browser plug-in that generates lots of meaningless web traffic to confuse anyone who might be collecting data about your web traffic, including your internet service provider (ISP). It visits random sites from a list you approve and randomly clicks around in the background while you let it run. Because it is constantly visiting sites—and it visits far more than you can on your own—its traffic will generally overwhelm your own traffic and disguise your actions.

* The Random Shopper bot was a computer program written by Darius Kazemi that would go to Amazon and randomly buy things. It had a budget, but besides that, it could order anything. This makes it very hard for Amazon to figure out what Darius really likes. Adding in misleading information is difficult as a strategy, though, because you really need to overwhelm the real information with fake. That's a lot of work!

Assignment

Showing results based on our location is a common way that a little bit of data can help us. This assignment will let you see how things differ if your location changes. First, do a web search for something that interests you. It could be a news topic, a hobby, or a sport. Do a second search for “pizza delivery” and see what kind of places come up. Next, you should do that same search but as though you were in a different country. LocaBrowser.com is one tool that will do this, but if you search for “browse the web from another country,” you will find other websites that let you do this. Pick a country and do the same web search from there. How do the results differ? Do you find them less useful?

Resources

Random Shopper, Tumblr, <https://randomshopper.tumblr.com/>.

Rejoiner, “The Amazon Recommendations Secret to Selling More Online,” <http://rejoiner.com/resources/amazon-recommendations-secret-selling-online/>.

What Is My IP Address, <https://whatismyipaddress.com/>.



ONLINE TRACKING: YES, YOU'RE BEING FOLLOWED

LECTURE 5

The more fluent you are in how data collection works from a technology standpoint, the more nuanced and effective you can be in your privacy-related decision making.

This includes understanding the process of web tracking: how websites keep track of you, your actions, and the data you produce.

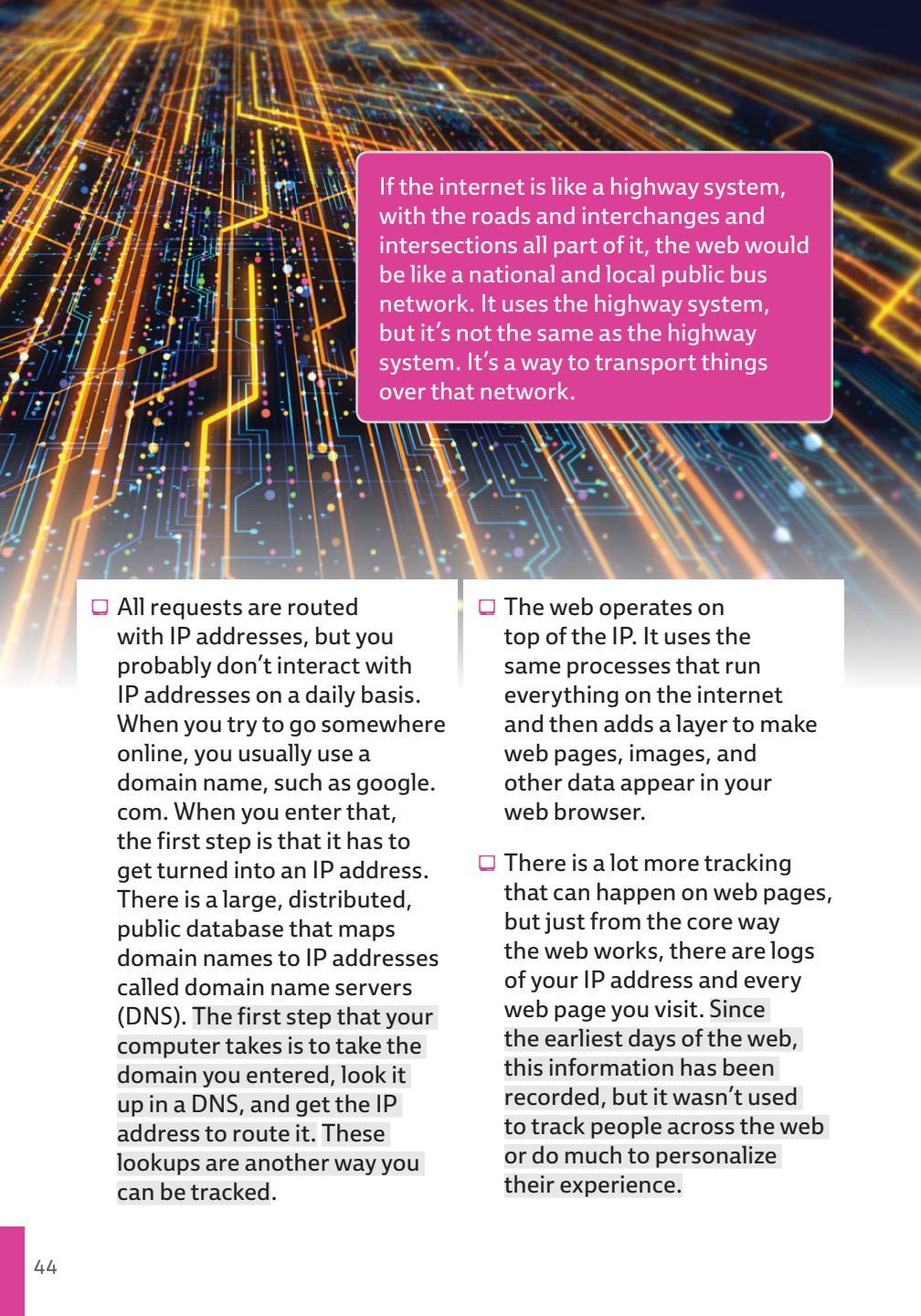


How the Web and the Internet Work

- The terms *web* and *internet* are often used interchangeably, but they're actually different things. The internet is the global network of computers. The web is a network of websites that operate on the internet.
- Information is transmitted over the internet on the web using protocols, which are basically rules about how to format, send, and receive information. The internet uses the Internet Protocol (IP), and the web uses the Hypertext Transfer Protocol (HTTP). An IP address is a unique address assigned to every computer and device that is connected to the internet; HTTP goes at the beginning of web addresses to tell your browser that you are using the Hypertext Transfer Protocol.*
- When you make a request for information online, there is a series of steps that gets followed. Say you're requesting a video so you can watch it on your device. Your computer puts together a request for this information, and it goes from your computer to your internet service provider (ISP), such as the cable company that brings internet to your house. From there, it may go to a few other servers** owned by your ISP. Then, it gets passed out to the internet backbone. From there, it connects to the server you are trying to reach. That server processes the request and sends a request back to your IP address in a reverse route. The exact path your request takes may vary each time.
- Servers keep a log of the requests they get and responses they send. This allows them to analyze their own performance. Your IP address is included in the log. This is the first way you can be tracked online.

* Even if you don't type the HTTP part, your browser fills it in because it's a necessary part of requesting web pages.

** Servers are just computers connected to the internet.



If the internet is like a highway system, with the roads and interchanges and intersections all part of it, the web would be like a national and local public bus network. It uses the highway system, but it's not the same as the highway system. It's a way to transport things over that network.

□ All requests are routed with IP addresses, but you probably don't interact with IP addresses on a daily basis. When you try to go somewhere online, you usually use a domain name, such as google.com. When you enter that, the first step is that it has to get turned into an IP address. There is a large, distributed, public database that maps domain names to IP addresses called domain name servers (DNS). The first step that your computer takes is to take the domain you entered, look it up in a DNS, and get the IP address to route it. These lookups are another way you can be tracked.

□ The web operates on top of the IP. It uses the same processes that run everything on the internet and then adds a layer to make web pages, images, and other data appear in your web browser.

□ There is a lot more tracking that can happen on web pages, but just from the core way the web works, there are logs of your IP address and every web page you visit. Since the earliest days of the web, this information has been recorded, but it wasn't used to track people across the web or do much to personalize their experience.

Tracking with Cookies and JavaScript

- ❑ On the web, there are a few technologies that have made their way into tracking infrastructure, though they were not originally intended to be used that way. One of these technologies is cookies, which are little pieces of code or identifiers that a website places on your computer.
 - ❑ There are lots of benign uses of cookies. For example, a website might use cookies to keep track of the fact that you are logged in or to remember your username for a login screen. It might keep track of what products you have viewed with a cookie. In fact, you can disable cookies in your browser, but most modern websites will not work without them.
 - ❑ Based on their original use, cookies could tell a website that you had visited it before. However, there are now more-modern cookies that can track you across many websites and aggregate that information to follow your visits more broadly.
 - ❑ JavaScript is another technology used in tracking. This is a programming language that is used for all modern interactive functionality on the web, but it is so powerful that it can be used to monitor everything you do, down to individual keystrokes you type in forms, even if you don't submit the data.
 - ❑ You've probably had the experience where you've been looking at a product and then ads for that product show up on other websites, even when you know there is no partnership between the two sites. This is called ad retargeting.
- Only the ad company tracks your browsing behavior across the web, not the individual sites, but it can be upsetting that any company is tracking you!

- For example, if Emily's Home Improvement uses ad retargeting, it partners with a web advertising company. The home improvement site puts a little code in their website that the ad company uses to track what products you have looked at. Then, other websites that partner with that ad company have a little code that tells the advertiser "find products that this person has looked at and show them." This uses a combination of JavaScript and cookies.
- In this case, the cookies that you agree to when you engaged with the website keep a note of products you have viewed. The ad company then retrieves those cookies to decide what ads to show you on other websites.

Tracking with Browser Information

- If you shop for a product on your home computer, the retargeted ads may appear on your phone or tablet. How do companies know it's you when you're using different devices? They do it through browser fingerprinting.
- A browser is the application you use to access the web. Internet Explorer, Firefox, Google Chrome, and Safari are all browsers. They know how to send requests for web pages, such as when you type in a search or a web address, and how to display the web pages' code in a nice way.
- Browser fingerprinting is a technology that can uniquely identify your browser by its characteristics, just like a fingerprint can uniquely identify you by its characteristics. How does this work?
- Hundreds of pieces of information about your system setup are available anytime you visit a web page, including which version of the operating system you're running, what fonts you have installed on your system, the dimensions of the window you have open, and what extensions are installed in your browser.

□ If an advertiser collects all of that information for every person, that information essentially becomes a fingerprint. It is not necessarily unique for every person, because two people could have the exact same system configuration, but it is unique in the vast majority of cases.

□ This allows advertisers to know that you are the same person visiting different websites, even if you don't have any cookies or other stored information. They can see that you are the person with that browser fingerprint, so they know what other pages you have visited. Essentially, they have a fingerprint for one of your fingers, and they can detect it in a bunch of places.

□ But how is this used to track you across devices? If you have an identifiable

configuration on your phone and on your computer, how do advertisers know to link those two profiles together? They can do this with account information.

□ If you are logged into an account on a website and visit it on your desktop computer, advertisers don't need to know any personal information about your account; they just need to know something like a username or user ID number that was logged in on that computer. That is often easy to identify. Then, if you log into the same account on your phone, the identifier tells advertisers that the same person owns these two fingerprints. Essentially, they now have prints for two of your fingers instead of just one.

About 80 percent of people can be uniquely identified by the configuration of their system, with no personal or otherwise identifying information.

- ❑ Advertisers store all of this information in a database so that when you visit a website, they grab the information provided about your system configuration and reference that against their database. This lets them know exactly who is visiting. You cannot block this system configuration information from being transmitted, so this is a technique that can be used on you regardless of what privacy settings or other privacy techniques you might implement. It is a very powerful way that your behavior can be tracked across the web, especially when these advertisers are large.
- ❑ Since many commercial websites, even personal blogs, have advertising, large advertisers may have information embedded on the majority of web pages that you visit. This means that not only do they know that you have come to a particular web page at a particular time, but they may actually be able to track the series of websites that you visit because every one of those pages has some of their code embedded on it. Even if they miss a few of those websites because their advertising is not used there, they get a pretty thorough picture of your behavior on the web.***

Tracking by Your ISP

- ❑ In addition to all the tracking already discussed, you can also be tracked by your ISP, which knows all the websites you are going to because the ISP provides the internet connection to your house. Thus, any request that you make has to go through your ISP, which sees where that request is going to and can keep a log of it.
- ❑ Until fairly recently, there was not much that ISPs could do about that. They were not supposed to use it to

*** The Electronic Frontier Foundation runs a tool that will analyze your configuration information and let you know if your browser fingerprint is unique and therefore likely to identify you.

target ads to you, and in the United States, the Obama Administration introduced rules in 2015 to make this illegal. However, one of the first pieces of legislation passed under the Trump Administration allowed ISPs to use this data to target you with ads.

- These regulations direct what ISPs can collect and what they can share. If your ISP is able to see what you do online and with its service, the ISP may decide to use this only internally to target you with advertising, or it could sell that information to other people. The recent legislation allows its use with outside advertisers, but internally, ISPs have been using that information for a long time.

- Many people get their internet service through their cable providers, and for a long time, cable providers have been telling advertisers that they can quite specifically target people based on their interests. This is not just with web advertising; this is with the commercials you see on your television.

- By combining some of your web activity with your viewing activity, your ISP can create a profile of your interests. Then, your ISP can use this to show you different television ads than it shows your neighbor, who may be watching the same program. Now, ISPs can also legally analyze your web traffic to enhance those profiles, and they have permission to use this information with advertisers online.

So What Can You Do?

- If you want to block the monitoring of your browsing behavior, you have several options. The first is to install browser extensions that block much of this kind of cookie and tracking activity. There are many tracker blockers available,

one of which is Ghostery. You can install these in your browser, and then, when you go to a page, they block all the trackers that they know about. Many will also create a report for you so that you know what was blocked.

- ❑ If there is a website that needs these kinds of trackers and that you trust, you can allow for an exception so that that particular site can use them. Not only does this protect you from a lot of tracking and improve your privacy, but it also can increase the speed at which you browse pages online. That's because it prevents all kinds of code from running in your browser and stops lots of places from putting code onto the web page, so there is less data to load and less processing that's happening in the background that's not necessary for your web experience.
- ❑ If you do a quick search in your web browser's extensions library, you'll find lots of extensions to block tracking. One thing to keep in mind, though, is that some pages rely very heavily on these kinds of trackers and simply will not function if they are blocked. To deal with that, you can either add exceptions in the blockers to allow the trackers to be used on those pages or install a second browser and use that on the rare occasion that you need to go to a web page that requires the trackers.



- There are two main ways that your ISP can see what you're doing online. The first is that they can see the actual web pages you go to because they bring those web pages into your home.
- The second is that they know what pages you want to visit because they turn the domain name—the thing that ends in .com or .net—into the IP address, the number that the internet uses to send you to a web page. Looking up a domain on a DNS to get an IP address is a necessary step to get information online. Usually, your ISP has their own DNS, which means the ISP can log all the lookups done for you and then know what pages you are visiting. Those are two separate steps, and both are ways that you can be tracked.
- If you want to stop your ISP from tracking you, you need to stop your ISP from seeing information in both steps. To stop them from seeing your domain name lookups, you can have someone else do that for you. In the network settings on your computer, you have the option to specify the IP address for the DNS you will use.
- There are lots of free, open DNSs provided by independent entities. You can easily find these with a web search. When you put those in as the DNSs to use, your computer sends the domain names to those servers instead of your ISP to get the IP address. That then blocks your ISP from seeing where you're going based on that lookup.
- The next step is to stop your ISP from bringing those web pages to your device. One way around this is through the use of a virtual private network (VPN), which essentially hides your web traffic from anyone who might be looking.
- With a VPN, instead of your ISP fetching a web page, your VPN host does it. Your computer establishes a secure connection with your VPN server, and your requests always go from your computer to the VPN service provider. They are encrypted, as is the information that comes back, so your home ISP is unable to see any information about what page you visited or the content that was on that page.

- A VPN is a great way to increase your overall security online and to protect the privacy of your data. There are free VPN service providers, but they can see all of the pages you're visiting and make money by selling information about you.
- Paid VPNs are affordable and give you a lot of security and additional privacy.
- By using an alternative DNS and a VPN, your browsing activity will be totally hidden from your ISP, and you get the added benefit of keeping your information much more secure.

Assignment

Check out how much you are being tracked. On your home computer's web browser and/or on your phone, get an extension or app that monitors tracking. A good phone app is Privacy Pro by Disconnect, and a good browser app is Ghostery. You will likely find others if you search; just be sure to check their privacy policies to confirm they aren't tracking you as well. Then, spend a day online like you normally do. Check in at the end of that day and see how many times you were tracked. It will likely be hundreds of times!

Resources

Disconnect, <https://disconnect.me/>.

Ghostery, <https://www.ghostery.com/>.

Golbeck, "How to Keep Your Web Browsing Private," <https://www.psychologytoday.com/us/blog/your-online-secrets/201703/how-keep-your-web-browsing-private>.

Rubinking, "How (and Why) to Change Your DNS Server," <https://www.pcmag.com/review/364418/how-and-why-to-change-your-dns-server>



NOWHERE TO HIDE? PRIVACY UNDER SURVEILLANCE

LECTURE 6

We've come to expect surveillance through digital channels, but surveillance is also out in the world in many ways that we may not suspect. We know that our devices collect data about us. But what about our interactions when we are off of our devices and out in the world? Just how pervasive is that sort of surveillance?



Monitoring of Health-Related Data

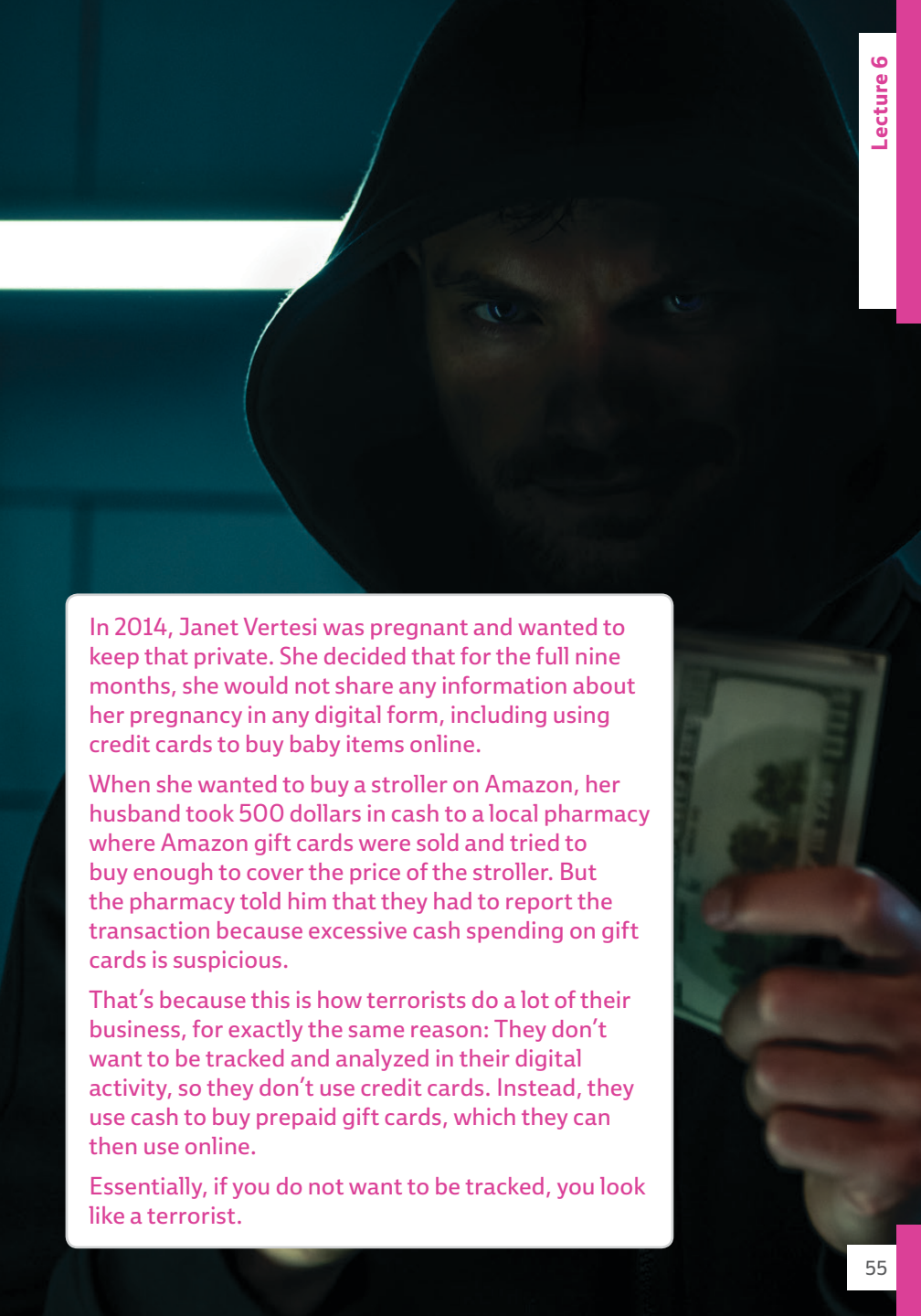
□ Companies are trying to collect more information about us with a veneer of consent, even though we may not know exactly what is going on behind the scenes. Health insurance companies, for example, offer some people discounts or gift cards if they link their fitness tracker with their insurance account. Of course, if you take a lot of steps, it makes sense that you might get rewarded for that. Auto insurance companies are similarly giving people tracking devices that can monitor their speed and driving habits. In exchange for the discount, people may give away their privacy in that domain.

□ But what about when we don't know? The *Houston Chronicle* shared a story in 2018 about a man who had sleep apnea and used a CPAP machine to help him breathe at night. These machines need replacement parts, such as filters and hoses, that insurance will pay for. When the man got a new machine, he registered it and opted out of receiving communication.

However, after the first night, he woke up to an email congratulating him on his use the night before. Later, he talked to someone at the company who mentioned that the device was working well at keeping his airway open. She knew that because she had a report of his usage.

□ This was something his old machine did, but that was recorded on a removable card that he would bring to his doctor's office. This machine, without his knowledge, was transmitting data about his usage. And it was sending it much more widely. It wasn't just going to his doctor; it was going to the company who made the machine and, to his shock, to his insurance company. And insurers use this data to deny coverage to patients who aren't using the machine enough.

□ Even with strong federal protections for health-related data, this type of monitoring seems to be legal when patients agree to the terms that come with their devices.

A person wearing a dark hoodie is shown from the chest up, looking directly at the camera with a serious expression. They are holding a stack of US dollar bills in their right hand. The background is dark and out of focus, with some horizontal light streaks.

In 2014, Janet Vertesi was pregnant and wanted to keep that private. She decided that for the full nine months, she would not share any information about her pregnancy in any digital form, including using credit cards to buy baby items online.

When she wanted to buy a stroller on Amazon, her husband took 500 dollars in cash to a local pharmacy where Amazon gift cards were sold and tried to buy enough to cover the price of the stroller. But the pharmacy told him that they had to report the transaction because excessive cash spending on gift cards is suspicious.

That's because this is how terrorists do a lot of their business, for exactly the same reason: They don't want to be tracked and analyzed in their digital activity, so they don't use credit cards. Instead, they use cash to buy prepaid gift cards, which they can then use online.

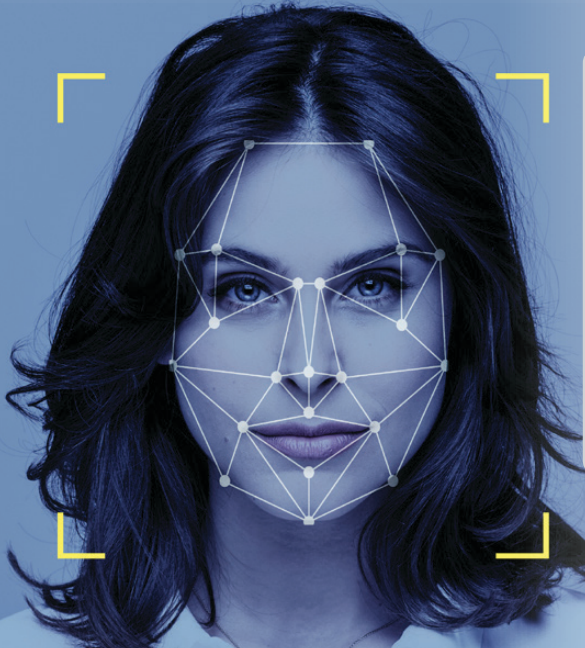
Essentially, if you do not want to be tracked, you look like a terrorist.

Facial Recognition Technology

- Outside the home, facial recognition technology is another space where real-world surveillance is becoming more sophisticated.
- Surveillance cameras are everywhere when we are moving around in public. Private businesses have them, and some municipalities have them. And devices like ATMs also have built-in cameras. As a result, our movements can often be tracked. But privacy is preserved in a way because so many people walk past these cameras. The images from the cameras are owned and controlled by many different people. As a result, it's difficult to aggregate all this together to follow a single person's movement.
- While that may change in the future as technology and integration develops, you only need to look to examples of police trying to track the movements of a victim or a suspect on cameras to see just how difficult this can be. It requires going into businesses and asking for copies of their video footage, which sometimes isn't working or sometimes is misdated or blurry. It requires watching hours of footage to try to identify exactly the right person and the time that he or she walked past and to reconcile that with what other cameras show. This, of course, makes things difficult for police, but for the average person who is just moving around, it means it's also very difficult for any large organization to keep track of our movements.
- Facial recognition algorithms can identify an individual by analyzing the pattern of the person's facial features. It's a technology that many large corporations are working on. Facebook has a good facial recognition algorithm. When you upload a picture, it automatically identifies the people who are in that photograph. However, not everyone has access to such a huge database of people's photos, so there are only a handful of companies with large and accurate facial recognition systems.

- Some of these companies, such as Amazon, are selling that technology to third parties. There has been a lot of controversy around this. The technology is not extremely accurate. It works much better for white men than it does for women and for people of color. That means that when errors are made, they are more likely to be made for those groups.
- This was highlighted in August 2018, when the American Civil Liberties Union

did an experiment using Amazon's facial recognition software. They compared 120 California lawmakers' images to a database of 25,000 mug shots. The algorithm incorrectly identified 28 state legislators as criminals even though none of them had ever been in jail and they were not the people matched in the mug shots. That is a pretty high error rate for an algorithm that is being deployed and used by police forces and other organizations.



In China, there is massive state surveillance that can indeed be used to track the movements of people on a large scale. Part of the way China is able to do that is with facial recognition technology.

- The way that this technology might be used makes this even more troubling. For example, there was a plan that has since been rolled back to link facial recognition technology and criminal databases with video doorbells. So when someone comes to your door, the video doorbell picks them up, runs their face against the set of databases, and can identify if someone with a criminal record is at your door. However, we know that there is a lot of inaccuracy in these algorithms, and they tend to be more inaccurate and make more mistakes on people with darker skin. This means it is likely to reinforce existing social biases.
- Furthermore, in neighborhoods where there is a higher density of people who have been in jail, it means that people's criminal records will be constantly at the forefront of everyone's mind. There are real social implications to doing this kind of thing even if the algorithms are right all the time.
- There is a lot of debate over the right way to use these algorithms. The inaccuracy and the potential for them to create a variety of social problems have led to bans on the use of facial recognition technology by government departments, including police agencies, in some cities. However, we are in the early days of this technology. It's possible that facial recognition may become more integrated into applications. It will require close monitoring if it is to be used in a fair way.
- In fact, even if the accuracy problems are solved, it's hard to say if there even is a fair way for this to be used. It would constitute a dramatic escalation in the way that people are monitored through their everyday movements. As one of the greatest threats to personal and civil liberties, facial recognition technology should drive the development of privacy legislation in the United States.

Tattoo Recognition

- Beyond facial recognition, technology exists to individually monitor people and their associations in other ways. Consider tattoo recognition. Facial recognition looks at the biometrics of your face to uniquely identify you. Tattoo recognition does a similar thing, scanning an image of the tattoo to distinguish it from any other.
- However, tattoos may be nearly identical between two people. If two people have the same flag or logo tattooed on their forearms, an algorithm may have a hard time telling them apart. Interestingly, the fact that they have the same tattoo may reveal that they are part of the same group, such as the same branch of the military or the same gang.*
- Data collection about tattoos is already quite advanced. The US National Institute of Standards and Technology provides government and law enforcement with a list of characteristics to note about tattoos, including type, location, color, and imagery.
- Law enforcement has long used tattoo imagery to identify gangs and members of hate groups, but tattoo recognition technology allows this to be carried to a new level. People on streets that are monitored with cameras, even existing surveillance systems, could have their tattoos automatically scanned, cross-referenced, and flagged as potentially gang-related. Essentially, an otherwise-anonymous person can be labeled as a gang member without any other action.

* Daniel Ramirez Medina, a 25-year-old immigrant who had been granted Dreamer status, was arrested in 2017. The government tried to strip him of his protected status, alleging that he was a gang member because of a tattoo he had. The tattoo actually was the name of the place his family was from in Mexico. Eventually, he was released, and a federal judge restored his status.

Advertising Kiosks

- ❑ But taking pictures of us and monitoring us are not just limited to identification for law enforcement purposes. There are now advertising kiosks that analyze your face as well.
- ❑ *The Wall Street Journal* reported that some shopping malls in South Korea had installed kiosks that have maps of the mall with lists of the stores, and each kiosk had a set of cameras and a motion detector. When someone came up to look at the map or browse the stores on the screen, those cameras and detectors used a facial recognition-type system to analyze the face of the person using the map.
- ❑ The malls were not trying to uniquely identify that person but rather to estimate the individual's gender and age. From there, the kiosk could drive the person to different stores or show him or her ads for other products. A young woman may see ads for something different than an older man would.
- ❑ This kind of processing respects privacy more than facial recognition does, but it is still invasive. It blurs the line between surveillance cameras that we have become somewhat used to—that monitor us in stores, presumably for public safety and to deter theft—and facial recognition technology that is monitoring and recording our movements as unique, identifiable people.
- ❑ When we are in public spaces, we know that we can be seen by other people who are there. And we know that we can be monitored in different ways. But we may not expect that the way we look, act, or move through those spaces will result in personalized advertising directed specifically toward us.
- ❑ Our reactions to this kind of technology should also consider how our data is handled. For example, in the kiosk situation, what is being stored? Is a person's age and gender being recorded or just used in the moment?

Could the kiosk owner analyze the demographics of people who used it? Are copies of people's photos being stored? Are they being shared with third parties?

- ❑ The fact is that when we walk up to a kiosk like this, we generally have no idea that a camera is present or that it is finding out information about us. Because of their privacy laws, a system like this is unlikely to be able to operate in Europe. Collecting this kind of personal data about a person would require explicit consent

and obvious transparency. And because European laws require an explicit opt-in, people would have to essentially push a button that says they are willing to have their picture taken and personal information analyzed in order to show them ads. That really defeats the purpose of passively analyzing people with a system like this.

- ❑ This kind of surveillance in the world highlights the need for legislation that will clarify what kind of privacy we should be able to expect and what kind of monitoring we can avoid.

So What Can You Do?

- ❑ There are ways to avoid being tracked digitally, but actually doing so can be almost impossibly difficult. So in terms of digital surveillance, it's really important to think about your comfort zone and what sort of effort you want to expend.
- ❑ Offline surveillance is even more difficult to control, because we so often don't know when we are being watched and to what end. Some surveillance in public is inevitable and has benefits

for public safety and security, but too much can threaten individual liberty and freedoms. The difficulty of analyzing that data has protected most of us from the most troubling consequences so far, but the technology and the algorithms are improving every day.

- ❑ This is a case where individual efforts at control might not be very effective. Surveillance and its consequences can only really be controlled through regulation and policy.

- If you feel strongly about surveillance and its impact on you, get to know the privacy laws that are in place in your country or community and become active in trying to improve those laws. Guidelines that bestow rights

on each of us to decide how much we share about ourselves, especially with profit-driven surveillance systems, are likely our best hope for a future with less monitoring, but we have far to go before those structures are in place.

Assignment

How often are you tracked out in the world? Next time you go out, make a conscious effort to tally all the cameras you pass. Include traffic cameras; surveillance cameras outside buildings, at gas stations, and in ATMs; cameras in stores; toll-collection cameras; and any other recording device you come across. Even keep an eye out for video-enabled doorbells. Search hard for them and see how many you can find. How much of your movement can happen without recording of some sort or another? It's probably less than you originally thought.

Resources

Allen, "With Connected CPAP Machines, Insurers Make 'You Snooze, You Lose' Literally True," <https://www.houstonchronicle.com/business/article/With-connected-CPAP-machines-insurers-make-you-13411742.php>.

Goldstein, "Meet the Woman Who Did Everything in Her Power to Hide Her Pregnancy from Big Data," <https://thinkprogress.org/meet-the-woman-who-did-everything-in-her-power-to-hide-her-pregnancy-from-big-data-80070cf6edd2/>.

Madrigal, "The Mysterious Printer Code That Could Have Led the FBI to Reality Winner," <https://www.theatlantic.com/technology/archive/2017/06/the-mysterious-printer-code-that-could-have-led-the-fbi-to-reality-winner/529350/>.



CONSENT: THE HEART OF PRIVACY CONTROL

LECTURE 7

Consent is a critical element of the conversation around our personal data and how it is used. If you have a thorough understanding of what data is being collected about you, how it's being used, with whom it is being shared, what decisions will be made with it, and you're OK with all of that, then there is no problem. But to give true consent requires all of those things, and today, the biggest problem surrounding personal data is a lack of attention to the part of consent that requires information.



The Ethics of Consent

- ❑ Consent is something that concerns many aspects of life, but in the world of personal data, there's a long history of consent being abused and people suffering as a result.
 - ❑ The ethics around consent are at the heart of how many researchers must operate in their work. Those same standards are not applied in industry. Companies don't think about consent in the same way academic researchers have to, but academic researchers came to the current consent systems as a result of some serious ethical violations and questionable practices.*
 - ❑ Now there is a nearly universal institution of informed consent for human subjects participating in any type of experiment. Researchers in government and academia must all have institutional review boards (IRBs), which are ethics committees that review every experiment that has human participants.
 - ❑ Participants need to sign a form that indicates their consent to participate before they begin participating. The form includes the purpose of the experiment, the ways participants' private data and identity will be protected, their right to stop participating, and other details.
- Privacy is not keeping things secret; it is deciding who to share what information with, at what time, and in what context.

* Up until the middle to late 20th century, all kinds of terrible, unethical things were done in medical research. Drugs, treatments, and vaccines were tested on orphans, the mentally ill, and prisoners all the time. People were unwillingly forced into experiments to advance medical science. Some of these led to good outcomes in the end, but at a serious ethical cost. However, many experiments were poorly conceived, were bad science, and led to a lot of unnecessary suffering.



□ This applies across the board, from something as simple as a survey to something as complicated as administering experimental cancer treatments. Obviously, the level of review from the ethics committee varies based on the potential risk that participants face, but you are not allowed to conduct research without

this process, nor will work be published without this approval and review.

□ Industry is not constrained by these rules. Some companies have their own IRBs, but most do not. Yet they are conducting similar research with similar risks—and doing it without giving experimental subjects any information or obtaining consent first.

Facebook's Social Contagion Experiment

□ One stark example is Facebook's social contagion experiment. Ever since Facebook became popular, researchers have studied the impact of the platform on people's happiness and self-esteem. A number of studies found

that people are less happy when they use Facebook. Obviously, Facebook doesn't think that it is creating a service that makes people unhappy and wanted to run its own study to show that its service did not make people sad.

- ❑ First, Facebook researchers used an algorithm to determine whether people's posts were happy or sad. Then, they came up with an experiment where a large number of Facebook users were separated into groups. One group would only see posts that ranged from neutral to happy; all the posts that were sad were excluded. The other group would only see posts that were neutral to sad; the happy posts were excluded.
- ❑ Facebook researchers wanted to measure whether people in the happy group posted more happy things and people in the sad group posted more sad things. If so, they would conclude that Facebook didn't make people sad but rather that people's moods were influenced by how their friends were feeling.
- ❑ Aside from the ethics of this research, this is a very poor experiment. If your friends are all sad because their pets are dying, their parents are sick, they have diseases, and they failed their classes, you probably aren't going to post lots of happy things, even if you are happy and your life is going great. Similarly, if all of your friends are celebrating and going through good times, you might be more cautious about posting something sad.
- ❑ You can't actually measure whether people are happy or sad based on what they are sharing because there are many things that influence what people share beyond their moods. But the Facebook study doesn't account for this; researchers conclude that if you are posting sad things, it's because you "caught" sadness from your friends, and the same goes if you're posting happy things.
- ❑ This experiment is also highly unethical. An academic researcher would have to submit a long proposal to the IRB to review this for ethics. Everyone who participated would have to be informed that they were being put in the experiment, and they would also have to consent to give their data.

- The vast majority of Americans are on Facebook, so it has a very wide coverage of every type of person in the country. This means that there are people who were on Facebook during the study who are severely depressed and suicidal. Is it ethical to manipulate the feeds of suicidal people to show them only sad things from their friends? Couldn't that have a real psychological impact?
- The risks that come from doing controlled psychological experiments on people are not minimal, and an ethical study needs to consider the psychological state of the people participating. This was not something that Facebook considered in its experiment.**
- And many experiments like this one are run often and every day, usually undetected by the people who are unwitting or unwilling participants. And there's not much we can do about it.

Transparency and Control

- Transparency means being clear about what is being done with users' data and how their experience is being controlled. Transparency generally involves clearly telling people what is being done with their data, who it is being shared with, and how it is impacting their experience. Ideally, privacy policies would cover this, but they tend to be written either with a lot of legal language that is hard to understand or with a lot of vague language—and sometimes both.

** Facebook eventually published an article on this experiment, and the journal in which it was published was inundated with ethical concerns from other researchers. It was forced to publish a response and to add a caveat about the ethics of this project to the published article.

- For example, it's not uncommon for these policies to say that your data will be used for a few specific things and then "research." This is a very broad term that could mean anything from analyzing which friends you want to see news from first to running a full-fledged psychological experiment. It also doesn't explain when those things might happen or how they're happening.
- It's also common for privacy policies to talk about data being shared with "third-party business partners." Those could be content distribution networks,^{***} which just provide additional server capacity designed to help host the content, but it can also include advertisers or entirely different companies. The language doesn't provide transparency because you can't read it and understand what is actually going on.
- One possible response to this might be that if you don't like the possibilities, then just don't use the platform. But in many cases, abandoning the platform is not an honest alternative. Even if you thought it was, a person can't meaningfully decide whether or not to use a platform when there is no transparency about what is happening with people's data or how it is being used in the platform. Thus, transparency is a critical element toward obtaining or assuming real consent.
- Control is the other element. Control means giving people the right to opt in or out or change their preferences.
- In this case, people have a choice about whether their data is used in certain ways. Social media platforms tend to give users some control about who can see information they post. These options may be simple or complex. For example, Twitter and Instagram allow you to essentially have a public or private account; Facebook has much more sophisticated controls, which let you choose individual people who are allowed or prohibited from seeing your posts.

^{***} Essentially, a content distribution network is a server rented by the social media company.

Knowing where you stand on the privacy spectrum is only part of the battle. Companies and people who use your data need to know where you stand, too—and respect it. Ultimately, respecting privacy means respecting people’s right to consent or withhold consent over how their personal information is used. And to truly give consent, you need to know a lot about what’s happening with your data.

- ❑ Unfortunately, these platforms give very little control to people regarding what companies can see their data when it is provided by the platform. The back-end data sharing remains mysterious. For example, it’s hard to know which of your Facebook posts or data derived from them can be seen by advertisers. That derived data is important to consider as well.
- ❑ There are laws that protect people’s data in some areas. For example,

in the US, the Health Insurance Portability and Accountability Act (HIPAA) contains a privacy law that controls who can see health information about you. Overall, getting a few signatures allows you to identify people or businesses who are allowed to ask for your health information, and it cannot be shared with anyone else. We provide this kind of consent all the time, and the process for doing so is relatively simple and straightforward.

❑ Does our online data deserve similar protection? This is a really hard question, because our online data is not really different than our general life data. But there are elements of transparency and control that can be easily and reasonably introduced to online data. For example, allowing us to determine which third parties have access

to our data from social media platforms would be relatively straightforward. There may be general cases we would always consent to, such as having a server company receive copies of our data because the company helps a platform work. In general, though, this is an area that requires much more attention and consideration.

Questions of Law

❑ Why is it that we have robust mechanisms in place in areas like health care, where people have to explicitly consent to their data being shared, but there are not even lightweight legal requirements in place for consent or transparency around social media data, which can also be very sensitive?

❑ In the United States and in other countries, the law says that once you share information with a third party, you no longer have a right to privacy over that information. In pre-internet days, this was a fairly reasonable law. But in the

age of the internet, this has come to mean that if you post information on a social media platform, even if it is covered by privacy settings and only intended for a small number of people, you've given up your right to privacy over that information because you have shared it with a third party—the social media platform itself.

❑ You also don't have a right to know what information companies have about you, nor do you have a right to have that data deleted or corrected if it's wrong, except in a few explicit cases, such as with credit reporting.

- The result is that in the United States, tech companies can basically do whatever they want with people's data and face very few legal repercussions. Most of the legal trouble that companies have gotten into surrounding privacy have come because they violated their own terms of service that everyone agrees to when they use the platform, not because they have actually violated laws governing the protection of information.
- Contrast this with Europe, where the fundamental idea is that people own data about themselves. Even if you share information with a third party, data about you belongs to you, and you have legal control over it. This has been the core idea of European law for quite a while, and most recently it was codified in a new way with the EU General Data Protection Regulation, which went into effect in May 2018.
- This law codified some basic rights for European citizens that Americans don't have. Basically, users are in control of their own data and are allowed to know what people have and what they're doing with it. The law also requires companies to disclose when and what they're doing with respect to personal data and tracking.
- It is unclear if this law is exactly the right way to handle privacy, transparency, and consent. However, the core idea that we own data about ourselves and should have control over it is one that many people would likely think is inherently a right we should have. That general legal idea would likely bring us to a place with much better transparency and clearer mechanisms for consent.

So What Can You Do?

- You can't possibly read, understand, and consent to all the terms of use and data policies that govern you on the web. It would take a huge amount of time just to read them, and even if you did, they may not clearly describe what the policies are.

□ But you can start by finding more efficient ways to understand the terms you are agreeing to. A long-running service called Terms of Service; Didn't

Read (ToS;DR) assigns an easy-to-understand letter grade to sites and, in a short bulleted list, highlights the good, concerning, and bad elements in their policies.

Assignment

Pick a website you use fairly often and where you have an account, such as your bank, a social media platform, or a news site. Check to see how transparent it is. Make a list of who you think your data can be shared with. Can the company share it with its business partners? How about its advertisers? Government and law enforcement? Any third party? Now review the terms of service or privacy policy and check your answers. Were you right—or, even more interestingly, was it difficult to tell?

Resources

Centers for Disease Control and Prevention, “U.S. Public Health Service Syphilis Study at Tuskegee,” <https://www.cdc.gov/tuskegee/timeline.htm>.

McLeod, “The Milgram Shock Experiment,” <https://www.simplypsychology.org/milgram.html>.

Meyer, “Everything We Know about Facebook’s Secret Mood Manipulation Experiment,” <https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/>.



DATA SCANDALS AND THE LESSONS THEY TEACH

LECTURE 8

There have been many scandals related to overcollection, poor protection, and manipulation of personal data. A few prominent examples are the Cambridge Analytica scandal, the failure of Google Buzz, employer social media spying, and the Ashley Madison hack.



The Cambridge Analytica Scandal

- Cambridge Analytica was a British-based company that worked primarily with political campaigns to profile people. Using artificial intelligence, the company gathered insights about the personality traits of these people and then used these insights to target them with political messages that were most likely to get them to behave in the way that the campaign wanted them to.
- During the 2016 US presidential primaries, news emerged that Ted Cruz was using Cambridge Analytica technology in his campaign. The idea that a company was psychologically profiling people and using that data to target them for political purposes was disturbing to many people. But there was something even more troubling: We knew that Cambridge Analytica had detailed personality profiles of people, but it was unclear how they obtained the data to build those profiles.
- In 2018, former Cambridge Analytica employee Christopher Wylie revealed that the firm had used Facebook to collect data on tens of millions of people without their knowledge and in violation of Facebook's privacy guidelines.
- Here's how they got the data. Cambridge scientist Aleksandr Kogan built a Facebook app that allowed people to take a quiz that would reveal their personality profile. The people who installed the app were informed that their answers would be used for an academic research project. When they installed the app, that app was able to access all of their Facebook data and the data of all their friends. People knew that their own data was being collected but not that their friends' data was being harvested. The friends were given no opportunity to consent or opt out; they weren't even aware of what was happening. So far, this is all permissible under Facebook's guidelines.

- ❑ Furthermore, the people who installed the app believed that it was only part of an academic experiment, not that the information was being turned over to a commercial entity. This is still OK under Facebook's guidelines, but it is a violation of the informed consent procedures commonly accepted and used by researchers.
- ❑ Next, Cambridge Analytica stored all of this data—information about the people who installed the app and the data about their friends. This step was the main violation of Facebook's data guidelines.
- ❑ Once they had the data and combined it with information from other sources, Cambridge Analytica was able to create detailed personality profiles of everyone in their data set. They then used this to target political messages that people received on a variety of platforms, including Facebook.
- ❑ When news of this broke, criticism started being directed at Facebook, which responded that Cambridge Analytica had obtained data from an unauthorized data breach. That makes it sound like someone hacked Facebook and took this information without them knowing, but that is not what happened.



Cambridge Analytica was involved in the Vote Leave campaign for Brexit and may have violated EU and British campaign laws in the process.

- ❑ Facebook just had no protections in place to keep people from storing the data that Facebook was handing over to them whenever they asked for it. So while this researcher at Cambridge did break the Facebook rules, Facebook also willingly gave him all the data that he eventually used with Cambridge Analytica.
- ❑ The impact that Cambridge Analytica's targeting had on the election is difficult to quantify because we cannot really tell who saw what or how impactful those ads were. However, this major data scandal prompted a lot of scrutiny of social media companies and questions about their stewardship of our data.

The Failure of Google Buzz

- ❑ While Cambridge Analytica was one of the biggest data scandals, it was hardly the first. One of the early privacy scandals in the social media space came with Google's first attempt to break into social media. Google Buzz was designed to be something of a competitor to Facebook. And it got off to perhaps the worst possible start.
- ❑ One of the challenges of starting a new social network is getting people on it. There are plenty of social networks that are better than Facebook in many ways. They respect your privacy more, have a nicer interface, and make it easier to share the content that you want with the people that you want. But the problem is that none of your friends are on those social networks, so there is very little value to them, even if they may be nicer to use than Facebook is.
- ❑ Google knew that friends make social media work, so they wanted their social network to be populated with lots of users right at the start. They realized that they already had many users who had Gmail accounts or accounts on other services that Google offered and could merge those into their social network.

Trust is a critical part of operating a social network, because you have access to so much information about people and their lives. When a social network or social media site breaches that trust, users are left with a decision about how much to risk by interacting with it.

Facebook has had a lot of scandals, but it has the benefit of being the holder of people's lists of friends and the majority of their social media posts. Thus, while people don't trust Facebook all that much, they face a big cost in leaving because all of their friends and content are there. However, when Google Buzz came along, people had nothing invested in it, so leaving it was very easy.

- ❑ This strategy alone is not a bad one. But Google also wanted to make sure that people who had accounts in the newly created Google Buzz also had friends, so Google decided to give users friends automatically by adding people who users emailed the most as their friends.
- ❑ This is a problematic choice from a social perspective. Of course, you may email someone often who is not your friend.
- ❑ On top of this bad decision, Google decided to make those automatically generated friend lists public. On one hand, this is typical of social networks;

friends lists are usually visible. However, by making a friend list visible to everyone, essentially Google was showing the world a list of the people that you email most.

- ❑ This created many problems. A therapist reported that he used email primarily to talk to his patients, so by generating his friend list from his emails and then making that public, the names of his patients were revealed to everyone, including to one another.
- ❑ The response to this was swift and harsh. And as a result, Google Buzz failed basically from the start.

Employer Social Media Spying

- ❑ One scandal that also came to a rather satisfying end was the practice some companies adopted of asking job applicants for the passwords to their social media and email accounts.
- ❑ During the recession of 2007 to 2009, as people lost their jobs and struggled to find new ones, companies were in a position to make a lot of demands of desperate people who needed income. This manifested in many ways, and some companies saw an opportunity to dig deeply into the personal lives of people who were applying.
- ❑ The first major report of this came in 2009, when the city of Bozeman, Montana, added a question for all job applicants that asked them to share “current personal or business websites, web pages, or memberships on any internet-based chat rooms, social clubs or forums, to include but not limited to: Facebook, Google, Yahoo, Youtube.com, Myspace, etc.,” along with their usernames, other login info, and even passwords. Even as the public outcry arose, other companies followed suit.
- ❑ If companies weren’t actively asking for passwords, they might require applicants to open their email and social media during an interview. Others required applicants to add someone in the organization as a friend who could monitor their posts, and still others were asking applicants or employees to set all their content to be public rather than private.

From the perspective of companies, as long as they can get access to personal data, they might as well collect more than they need, and if there’s future value in the data, it will be worth the small cost of storing it in the meantime.

But companies are not necessarily ethical or secure in their treatment of such information. Because cybersecurity laws are so lax, there is not a huge incentive for companies to thoroughly protect personal data.

- Companies saw this as an opportunity to dig deep into the backgrounds of their applicants in a profoundly unethical way. And because they could take advantage of the poor economic situation and people's desperation to find jobs, they got away with it for a while.
 - But fortunately, within a few months of news reports surfacing of these practices becoming more widespread, many US states began drafting legislation that forbid companies from these practices. Nearly half the states eventually passed laws to protect people.
 - The outrage and fast legislative response show that this practice was widely considered unethical and an abuse of power. But there are other reasons it is a bad idea.
 - First, it may reveal information about a candidate or employee that is illegal to use in hiring decisions. For example, employers cannot discriminate against pregnant women.
- In an interview, if a potential employer forces a woman to bring up her social media posts, which include posts about her pregnancy, and then doesn't hire her, she could reasonably argue that they used her pregnancy against her in an illegal way.
- In addition, if an employer asks for an applicant's password and the applicant sends it to the employer electronically, this creates a huge cybersecurity problem. By storing an unencrypted* password—storing it in plaintext—anyone can see and use that password. Of course, this is why the companies were asking for them; they wanted to log in and see what people were doing in their private accounts. But not only does this open up access to the person's account to anyone who gets ahold of that password, but it also means that if the system where that password is saved gets hacked, the password would be exposed. For sensitive systems like email and social media, this is a huge risk.

* Normally, companies who have a login for you only store the encrypted version of your password. That way, if their systems get hacked, people don't get your actual password.

- It also likely puts the companies in a position where they are out of compliance with regulations or policies that govern their own cybersecurity.** Collecting and storing passwords for applicants' social media and emails would not pass any test with respect to industry security standards.
- Fortunately, this problem now seems to be under control because of the legislation that was quickly passed to block the practice.

The Ashley Madison Hack

- One very dramatic hack of personal data happened with the website Ashley Madison, which is explicitly marketed as an adultery dating site.
- A huge problem with the website was that you could create a profile for someone else using just an email address. Anyone else who went to the site could then search for that person and find the profile you set up. And because Ashley Madison knew that a lot of people would want to delete their profiles—even if you set up an account on your own, you might decide that having a profile on a cheating website was a bad idea—the company charged people to take down their profiles. They reported bringing in 1.7 million dollars a year just from these fees.
- On top of that, it turns out that Ashley Madison was not actually deleting the data that was supposed to be deleted when people paid the fees!

** Many larger companies and government agencies have to comply with a variety of standard practices to maintain certifications that are necessary to their business. As part of this, they are audited for their cybersecurity practices.

- Because of the adulterous nature of the website and because of the exploitative data practices, Ashley Madison became the target of many groups who were concerned with their ethics. This all came to a head in 2015, when a group calling themselves the Impact Team breached the systems of Ashley Madison and downloaded all of their data.
- The Impact Team told Ashley Madison that they would post the data, exposing users' identities, unless the website shut down immediately. Ashley Madison refused to do so and denied that there was a massive data breach. The Impact Team called their bluff and released all of the user data—including full user profile information, user-to-user messages, street addresses, credit card numbers, and photos—onto the dark web.
- The humiliation and repercussions from an adultery website being hacked were far greater than it would have been with other domains. A number of suicides were tied to the hack. People and unscrupulous organizations combed through the data and used it to blackmail or threaten people in the data set. Companies searched for people who used their work emails to register for accounts. ***
- Class action lawsuits were filed against the company that were settled for millions of dollars, but the company still exists, with millions of followers.

So What Can You Do?

- If you've followed the steps outlined in other lectures, you're in pretty good shape. Limit the data you store online, delete frequently, and use good privacy and cybersecurity practices.

*** There were 1,200 Saudi Arabian (.sa) email addresses in the hacked Ashley Madison data, and in Saudi Arabia, adulterers can be punished with death.

- ❑ However, any system can be hacked, and when our data is in the hands of many organizations, we have to trust them to be good stewards. Not all companies live up to those responsibilities, though, especially when they can make money by spreading our data around.
- ❑ Many of the solutions to these problems are only possible with legal and regulatory changes. In the meantime, one best practice might be to maintain a good dose of skepticism about the security of your information. Even if you trust the company with whom you are sharing that information, that trust can be manipulated or broken in many ways. So share your information sparingly and wisely.

Assignment

Check to see if your data has been breached. You may have received a notification for some major hacks, but do you know the full extent? There are tools online that will search for your email in databases of hacked information. Have I Been Pwned (haveibeenpwned.com) is a legitimate and long-standing tool, but there are many others. Check for your email address in some services and see where you stand. Do the results surprise you?

Resources

Bhat, "Google Buzz," <http://stlr.org/2010/02/19/google-buzz-a-recap-of-the-controversy-and-the-current-legal-issues/>.

Chang, "The Facebook and Cambridge Analytica Scandal," <https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram>.

Lord, "A Timeline of the Ashley Madison Hack," <https://digitalguardian.com/blog/timeline-ashley-madison-hack>.



THE DARK WEB: WHERE PRIVACY RULES

LECTURE 9

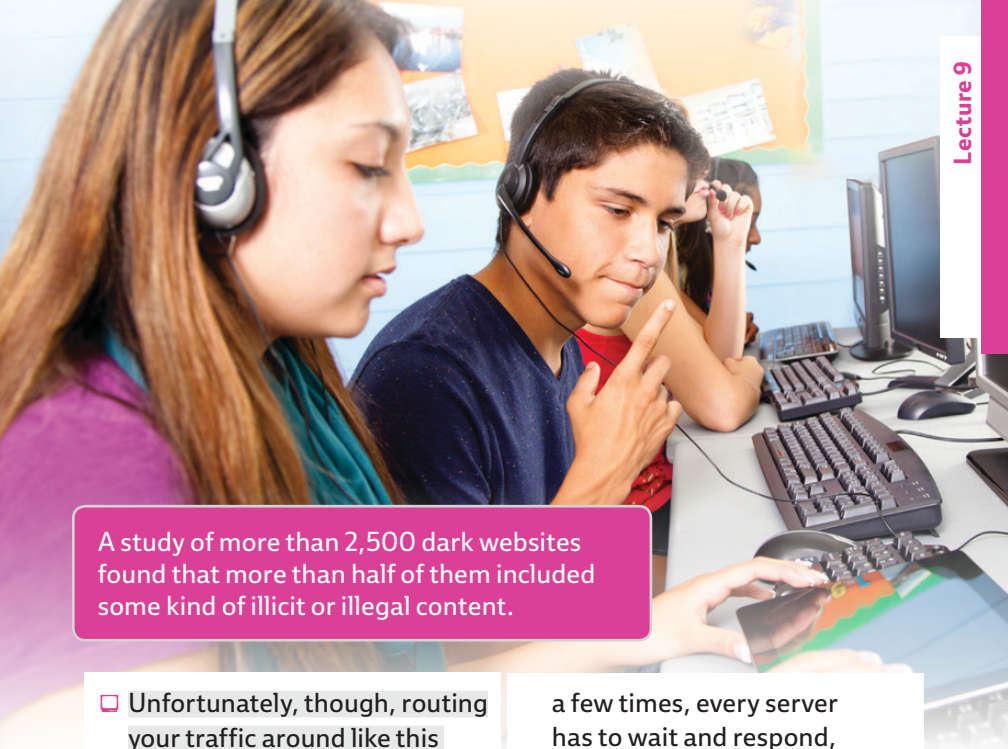
Although the dark web is known as a place where illegal activity happens on the internet—from drug and weapons dealing to trading software viruses—it is also a place where plenty of people go for legitimate reasons, including because they want privacy.



Using the Tor Browser

- ❑ The dark web is not accessible from regular web browsers and is not indexed by search engines. It uses the same technology as the web and operates with web browsers, but to get there, you need to be able to access that part of the web network. This is done using the Tor browser*—which you can download for free and use just like you would any web browser. The key differences are that it can access sites on the dark web, and it protects your web traffic from snooping.
- ❑ Tor is built on top of Firefox, so it looks and works mostly like the Firefox web browser. However, it's designed to protect your web browsing by routing it differently. Instead of connecting you directly to the web page you want to access, Tor routes your traffic through a series of intermediate servers so that no individual server knows where you are actually connecting from.
- ❑ Any particular server in the series of servers only knows the location of the server that came directly before it in the chain. Thus, even if your traffic were intercepted at one of the servers, no one would be able to track it back to your computer. Eventually, your request will reach the site you want to visit, which will then return the page to the last server who requested it. That server will pass it back in the chain, and this repeats until the page finally reaches you.
- ❑ This gives you a great deal of privacy with respect to your web searching habits, since no one can trace a request back to you. Your home IP address is only known by the first server in the chain. Servers on the Tor network do not log this information, so no one can piece together that chain to track a request back to you.

* Tor, which originated as an acronym for The Onion Router, was originally developed by the US Navy.



A study of more than 2,500 dark websites found that more than half of them included some kind of illicit or illegal content.

❑ Unfortunately, though, routing your traffic around like this dramatically slows down your web experience. If you try to go to Google from your home computer in a regular browser, you hardly notice a delay before the page appears on your screen. On a slow day, it may take one second before it appears. With the Tor browser, accessing that same website in the same way may take five or 10 seconds. Sometimes, it even fails, and you have to try to connect to the website again. That's because if you route your request around the world

a few times, every server has to wait and respond, and that really slows down what's happening. Ultimately, whether you think that kind of delay is acceptable or not comes down to personal preference.

❑ There are other inconveniences that come with this traffic routing. The website at the end may be using basic information about where you are coming from to determine what to show you. If it looks like you are coming from another country, some websites may not work.

- People may want this kind of protection simply if they are privacy-conscious, but it becomes more important if you live in places where you know that your web traffic is being monitored. In many countries who do this kind of monitoring, VPNs—virtual private networks that encrypt data coming from your computer—are banned, so traffic can't be hidden that way. Tor provides a way around this, allowing people in countries with oppressive governments and restrictive internet to access whatever sites they want while covering their tracks.
- Of course, people who are engaging in illegal activities also want this kind of privacy. But it is perfectly legal to use the Tor browser, and many people use it for legitimate purposes, so there's nothing wrong with downloading it and giving it a try!

Dark Web Anatomy

- The dark web does not use different technology from the regular web. The main way that you can tell the difference between dark websites and regular websites is that dark websites all end with the .onion top-level domain instead of ones you're familiar with, such as .com or .net. If you try to access a .onion website with your regular browser, your browser will just think that you have entered an incorrect address, and the browser will not be able to get to it. The Tor browser, on the other hand, can access these sites.
- The domain names of .onion sites look different than what you would expect on the regular web. Instead of being able to choose your own domain name, every dark web domain name is 16 characters followed by .onion. Just like with the regular web, anyone who is connected to the dark web can set up a server and host a website.

Some websites you're familiar with exist on the dark web. For example, Facebook is on the dark web as facebookcorewwi.onion.

- ❑ On the regular web, if you want a domain name, you need to register it with a domain name registration service. That maps your domain name to the IP address of the computer that hosts your website. A distributed database of these mappings is kept on domain name servers.
- ❑ On the dark web, there's a similar domain name service, but instead of just choosing the word that you want for your domain name, you basically pick from a list of all available 16-character strings. This means that almost every website on the dark web has a meaningless domain name that's just a bunch of letters and numbers followed by .onion. That means that it's pretty much impractical to memorize domain names on the dark web like you do on the regular web.
- ❑ Because domain names essentially can't be memorized, it would be useful to have good search engines for the dark web, but that's not the case. There are dark web search engines, but they are more like using a search engine from the 1990s on the regular web. The results are often irrelevant, lead to broken web pages, and are missing a lot of relevant information. This is not because there are no professionals building the search engines, but rather because things change at a much faster rate on the dark web.
- ❑ There is a lot of nefarious activity going on there. Popular sites will get attacked by hackers that bring the service down, or they are targeted and shut down by law enforcement because illegal activities are taking place there. Once they are gone, it's very easy for them to simply reopen with a new dark web domain. And because those domain names are not meaningful or memorized, it's not like they are losing important branding. But frequently changing domain names means that search engines can't rely on sites being in the same place for very long. Thus, many of the ways that you discover things on the dark web is by word of mouth.
- ❑ Websites on the dark web also look very outdated. They tend to have very simple interfaces and don't look as professional as regular websites, but the advantage of this is that they load very quickly.

Dark Web Activities

- ❑ In the context of your personal data, there are two relevant reasons to use the dark web: It's where personal data that has been stolen can be found, and it's a way to keep your activities more private.
- ❑ On the dark web, there are plenty of perfectly legitimate activities going on, including people playing games and having political debates. If you are discussing sensitive topics, you can do that anonymously in a way that cannot be tracked. In this way, the dark web embodies the ethics of the early web that focused on freedom of expression—and that freedom as a tool for improving people's lives.
- ❑ But the dark web is also a place where many illicit things happen. For example, you can find the full text of popular books, along with pirated content.
- ❑ One of the major activities that you can do on the dark web is to buy things, and you can buy pretty much anything—including stolen credit card numbers, stolen login information, drugs, guns, pornography, computer viruses, and the services of people who will help you do illegal things, such as hackers, currency traders, and hitmen. The marketplaces where these types of things happen look like low-rent versions of eBay.
- ❑ In addition to the anonymity offered by the dark web, the rise of bitcoin** and other cryptocurrencies has enabled these sorts of transactions to take place anonymously and securely. Transactions with cryptocurrencies are recorded in public ledgers maintained by volunteers. The transactions are anonymous, with each person identified only by a string of letters and numbers. And the data within them is encrypted, so it remains secret to everyone except the two people in the transaction.

** Bitcoin and cryptocurrencies are currencies that were invented; they are not tied to any government or company.

- ❑ Ideally, you can convert cryptocurrency into any other currency, but it is highly volatile, and whether or not conversion works reliably and safely is still up for debate. It often involves meeting strangers in parking lots to do the exchange. You can buy bitcoin and other cryptocurrencies with regular money, trade it on exchanges, and use it to buy things on the dark web.***
- ❑ In terms of personal information, you can buy bank account numbers and login and password information for bank accounts on the dark web. Using this sensitive personal information comes with a huge risk, but it can be had for a price—a relatively low one, in fact.****
- ❑ When there are data breaches of username and password information for large websites, that data tends to end up on the dark web. It is not as valuable as bank account information, but it can be used in many ways.
- ❑ Hacked personal information can also be aggregated, so there are places on the dark web where you can find a person's collected email addresses, login names, and hacked passwords, along with other information that may have been obtained legally, such as credit card numbers and Social Security numbers. All of this extremely sensitive information is available for a relatively low price to anyone who wants to buy it.

*** Cryptocurrency and the dark web marketplaces have evolved together: The marketplaces make cryptocurrencies more useful, and cryptocurrencies enable dark web transactions to take place securely and privately.

**** In 2017, Experian found that on the dark web, a Social Security number cost just one dollar, a credit card number with the code on the back was five dollars, a debit card number and the associated bank information was 15 dollars, and a driver's license was 20 dollars.

❑ If you have any kind of normal online presence, there is probably information about you for sale on the dark web. Unfortunately, there's not a whole lot that you can do about that. Because of all of the

privacy and security elements associated with the dark web, it's difficult to shut these repositories down. They just pop up someplace else and are not traceable to the individuals who are running them.

So What Can You Do?

❑ If you want to know what's on the dark web about you, you could get on the dark web yourself and start searching, but plenty of places like credit bureaus and credit card companies now offer dark web monitoring that looks for your personal information on the dark web and alerts you to it. If they find your credit card number or a password that you are still using, they can alert you that this is something to change to keep other accounts secure.

❑ However, these services come at a cost. Sometimes it's in random fees that you are charged, and often it means that you give up your right to sue the people monitoring you, even if they are the reason your information ended up on the dark web in the first place.

❑ The only real steps you can take to protect yourself are using good security practices. Using things like two-factor authentication will alert you if anyone tries to get into your accounts, and it makes it harder for people to access them. You can set up a credit freeze or monitoring with a credit bureau, but be sure to read the fine print.

❑ But beyond that, the dark web and the dark marketplaces of information that exist there are just an unfortunate reality of our modern digital life right now. Hackers are going to hack, and until there are stiffer penalties for companies to encourage them to do much more to protect our information, that information will fall into the hands of criminals.

Assignment

You may not want to get on the dark web itself, but try out the Tor browser. You can use it for any normal web browsing. You can download it at torproject.org. Try visiting a few websites you normally go to. Can you notice a slowdown in the browsing speed? Do you have difficulty accessing some sites? These are side effects of the privacy protection, and they highlight the trade-offs we often have to make when balancing privacy against convenience.

Resources

Bitcoin, <https://bitcoin.org/en/>.

Stack, "Here's How Much Your Personal Information Is Selling for on the Dark Web,"

<https://www.experian.com/blogs/ask-experian/heres-how-much-your-personal-information-is-selling-for-on-the-dark-web/>.

Tor, <https://www.torproject.org/>.



ALGORITHMIC BIAS: WHEN AI GETS IT WRONG

LECTURE 10

While algorithms can be eerily correct at predicting people's desires and behavior, they can also be incorrect, with troubling consequences. When algorithms go astray, it can often be traced to bias within them. As algorithms become more widely used outside the research lab, bias can have real and potentially damaging effects on people's lives. It is important to understand how this bias manifests, where it comes from, and how it can be addressed.



Bias in Algorithms

- Facial recognition algorithms can use a picture of your face—whether it’s posted online or captured in a surveillance system—to identify who you are by comparing it to a database of known faces. But such algorithms can be wrong; in fact, they’re wrong in biased ways.
- If you are a white male, these algorithms work very well for you. If you are a member of any other group—such as white women or men of any other race—they are less accurate.
- How does that happen? It’s not that there was any racism or sexism at play in the development of the algorithms; it’s that in order to work, these algorithms need to learn from a bunch of examples of known faces.
- A social media company might have photos for many of its users. People share lots of photos, and if they tag who is in them, that can be used to teach a facial recognition system who each person is and, on a larger scale, to learn which facial measurements are useful at identifying people. But if you don’t have access to that kind of data, you need to look to public databases of identified photos. One of the most common databases used for this is one of celebrity photos. There are many celebrities, and there are many pictures of them taken from many angles, making this a useful set of faces to learn from.
- But for this to work, celebrities need to be representative of the general population. In a way, they are. We all look sort of like celebrities; we all have the same face parts in roughly the same spots. But celebrities also look much different than most people. And depending on your race and gender, the variation within celebrities can be different.

- For example, famous white men tend to have more variation in their age, weight, and range of appearance than famous women. While it is relatively common to see male celebrities over 50, there are many fewer women of that age who are celebrities. Female celebrities are also expected to conform to much more specific beauty standards. Male celebrities tend to have more varied body types than female celebrities do.
- This is not to say that there is no discrimination based on celebrity men's appearances but rather that male celebrities tend to have much more variation than female celebrities. Thus, identifying men is easier because the algorithm has learned from a more diverse set of male examples.
- This shows that algorithms can be less accurate for certain groups almost by accident, because the data the algorithms learn from is not representative across all groups. The bias is not embedded in the algorithm intentionally but is a result of bias in the data—and even the bias in the data is more accidental than intentional. But sometimes bias makes its way into algorithms in a more overt and malicious way.
- Consider lending, whether for credit cards or mortgages or bank loans. Not long ago, lending was legally biased. Women were required to have their husbands on their credit cards. With a restriction like this, women were not able to independently establish credit ratings in the same way that men could. Even after that rule changed, it would take decades for women to establish the same credit histories as men and appear as reliable. This kind of social bias against a group has long-lasting impacts that appear in the data.
- And just because a bias is no longer legal, that doesn't mean it disappears. Human decision making is often biased in racist and sexist ways. While some of these biases may be explicit, some of them are unconscious and unintentional.

Facial recognition algorithms can be correct nearly 100 percent of the time for white men but only about two-thirds of the time for nonwhite women.

- There is evidence that mortgage rate decisions by lenders are influenced by human bias—often racial bias. Perhaps higher rates are offered to minorities, even if they have good credit. And when we build an algorithm that learns from the data created by these human decisions, it learns a similar bias.
- One example of this took place in the lead-up to the mortgage crisis. The city of Los Angeles did an investigation and found that among people with the exact same credit history, African Americans were steered into subprime mortgages one and a half times as often as white people, and Hispanics were sent to subprime mortgages twice as often. Those decisions were made by an algorithm.
- Certainly, the bank knew it was illegal to set mortgage rates based on race, and the race of the applicants was not something that the algorithm was told. So how did the algorithm end up making racist decisions?

- It learned from human data, which has some of that bias in it, because humans are biased. From that data, it recognized that certain people were sent to subprime mortgages more often. And while it couldn't see race, it could see zip code, which maps closely to race in much of the US. Thus, even though the factor that was

being discriminated against was excluded, artificial intelligence (AI) was smart enough to find other features that allowed it to mimic that bias. Essentially, by discriminating based on zip code, the algorithm was being more "accurate" because it was more closely reflecting human decision making.

Inscrutability of Algorithms

- The problems of bias are also exacerbated by the inscrutability of algorithms. They can produce an answer but rarely produce an explanation to accompany it. Often, that's because algorithms cannot explain how they made a decision. The internal math is impenetrable and complex, so there is no way to know how they came up with an answer.
- For example, in *Weapons of Math Destruction*, Cathy O'Neil describes an algorithm that was used in Washington DC to determine whether or not teachers were retained. She tells the story of one excellent teacher who was loved by students, parents,

and principals. The school system implemented an algorithm to rank teachers, and the school district fired the lowest-scoring five percent.

- This particular teacher was in that group, much to everyone's surprise. She talked to the school about what happened, and the answer she got back was essentially "the algorithm says so." It's known that these algorithms vary widely in their accuracy, with teachers scoring at the top one year and in the bottom the next—with no change in classroom techniques or strategies. Yet she was not given the opportunity to challenge those results and was let go.

- This teacher was only unemployed for a few days before getting rehired in a wealthy suburban school district outside Washington DC. This outcome is exactly what we want to avoid: An excellent teacher in an underresourced urban school system who wants to stay there ends up being fired and bringing her skills to a wealthy suburban school system. She's exactly the kind of person that districts want to retain, yet an algorithm—which was probably flawed—gave an answer that people decided to rely on without question to make a decision.
- When there is no insight into the algorithm's decision process like this, it can be difficult or even impossible to detect bias. Was this algorithm systematically biased in some way? Did it simply make a random error? Is there some other important factor at play that we don't know? It's impossible to know.

So What Can You Do?

- The combination of algorithmic errors that go unquestioned and algorithmic bias that is hidden by the veneer of objectivity that comes with AI means that there is a lot of social bad that can arise from these kinds of technological approaches being implemented without question, without checks, and without understanding.
- Unfortunately, it is common for algorithm builders to offer a naive solution to the problem: If an algorithm is sexist, for example, we could just drop out indicators of gender! But this does not work. Countless examples show that AI is smart enough to pick up on a person's gender from other clues. Hiding the obvious indicators of a trait may be good at hiding the bias, but the bias usually will still be there.
- There are a few ways to approach the use of these algorithms to make things better, including algorithmic explainability and auditability as well as diversity in algorithmic design, though the problem is large and complex.

- ❑ Part of the difficulty that comes with using AI is that these algorithms are very difficult to understand. Under the hood, there are many layers of complexity that lose human meaning very quickly. Even though we can output all the numbers and data that an algorithm uses, we can't understand what it means or how it came to a decision.
- ❑ In fact, in the 1990s and early 2000s, computer scientists spent about a decade trying to come up with a way to explain how these algorithms made their decisions, but ultimately they gave up because it was too hard. Thus, if someone asks if an algorithm is discriminating based on gender, it's very hard to answer based on what the model looks like. You can't just peek under the hood and see that it considers gender in a specific way.
- ❑ One important area of research is trying to figure out how we might be able to better explain what's going on inside algorithms. It's a place where we hope to make fast progress because as these algorithms are more widely used, it will become important to explain what they are doing. This is one of the spaces where we're likely to see some of the earliest regulation over the use of AI.
- ❑ This may come from using different types of algorithms. Some are much easier to understand than others. For example, one class of algorithms, called decision trees,* aren't always as accurate as some of the more complex algorithms, but if companies are required to explain their algorithms' choices, they may end up sacrificing some accuracy for more interpretable results.

* With decision trees, you start with a question on the top level, such as "Is it raining?" Depending on the answer, you will go to another branch, which will ask a different question, such as "Is the temperature 80 degrees or above?" Eventually, the branching stops with a decision like "Today is a good day to visit the beach" or "It's a bad beach day."

- There's also research going on that hopes to better understand what's happening inside the more complicated algorithms. Technology has advanced a lot since the last time researchers worked on making these explainable, so there is some hope that there could be progress there.
- Another approach—called auditing—is to analyze the output of the algorithms rather than analyzing what's going on inside them. The processes for this can be complex, but essentially it involves sending a bunch of different examples to the algorithm, seeing what kind of output is generated, and then evaluating that for bias.
- Just how to do this auditing is a difficult problem, and it's an active area of research among AI scientists. Because of the ethical and legal context around it, this is also a space that we are likely to see a good amount of progress in the coming years.
- With AI being used to hire and fire, make lending decisions, and potentially determine insurance coverage, health-care access, and other financial decisions, there will certainly be accusations of unfairness and lawsuits. Users of the tech will need some defense, and comprehensive audits are a way to show that care has been taken to screen the algorithms for problems. They are also important before things go as far as a lawsuit for companies to know how well their algorithms are working and where any unfairness lies before they put them to use.
- In addition, companies will want to avoid building bias into their systems in the first place. This means having diverse teams working on creating the algorithms. Some of the bias issues that arise come from a lack of knowledge on the part of the algorithms' creators. The bias is not necessarily malicious, but a product of ignorance or naivete.

- On the personal data side, these issues are also relevant. More diverse teams lead to better decisions about data privacy as well as to better insight about fairness and bias—one of the biggest challenges as algorithms that use our personal data become more widely deployed. As average users of technology, we don't have a lot of control over this. But it is an issue worth paying attention to because ultimately there is likely to be regulation over these issues that will directly impact our lives.
-

Assignment

Many search engines have an auto-complete function on their home page. You might type in “chocolate cake” and it will offer you options like “chocolate cake recipe” and “chocolate cake from scratch.” Try searches for the same phrase for different groups. For example, try typing in “is my daughter” versus “is my son” or “why are men” versus “why are women.” Look at the different suggestions offered and compare them between the two groups. Are the differences reflecting stereotypical biases? How do you think they got there?

Resources

Dastin, “Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women,” <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.

Jackson, “Facial-Recognition Technology Flagged 26 California Lawmakers as Criminals,” <https://www.cnn.com/2019/08/13/tech/facial-recognition-body-cameras-bill-california-trnd/index.html>.

O'Neil, *Weapons of Math Destruction*.



PRIVACY ON THE GLOBAL STAGE

LECTURE 11

Around the world, there are very different approaches to how personal information is used, protected, and shared. While the default in this course has been viewing things from the perspective of the United States, two very different examples are Europe and China. Europe is leading the world with a strong, comprehensive privacy law that grants citizens protections that many in the US envy, whereas China is carrying the collection and use of data on citizens to a level that many would consider dystopian.



United States

- ❑ In the United States, private communications are protected. The Fourth Amendment gives citizens protections from unreasonable search and seizure, which prevents the government from intercepting phone calls or mail or otherwise spying on you without a reason.
- ❑ There are other legal protections that help keep your information private from the government and from third parties. For example, a person is not allowed to read your mail or come into your house. When determining when and how data can be collected about an individual, what's

discussed is whether someone has a reasonable expectation of privacy.

- ❑ When we share communications with third parties, though, we lose our expectation of privacy. In the modern context, that means that if you post something on social media—even if you have privacy settings that make it visible to only a few people—you cannot expect that it will be kept private. You have shared it with the social media platform in the process, and now it essentially owns that data about you. The company is not subject to the same rules that the government is.

There are not many rules that restrict what third parties can do with your data. This comes from the 1979 Supreme Court case *Smith v. Maryland*, in which the court upheld the third-party doctrine, which basically says that once you involve a third party in communication, you lose your legal privacy protections.

- Most of the restrictions over what companies can do with your data come from the companies themselves. They must abide by the terms of service that you agree to when signing up, but those can change over time. When companies violate their own terms, people may sue. Settlements may result in simply changing the terms of service to allow the offending activity, or the government may get involved.
- For example, in 2011, the Federal Trade Commission (FTC) investigated Facebook for telling users their information would be kept private while making it public in a variety of ways. To settle the charges, Facebook agreed to a consent decree that regulated what Facebook could do going forward. It said that Facebook had to get consent before sharing data with third parties, develop a comprehensive privacy program, and have privacy audits every two to three years.
- After the Cambridge Analytica scandal broke, regulators dug into how apps were able to collect information about millions of unsuspecting Facebook users. An investigation that stretched over two years found that Facebook had violated the decree in many ways, sharing data that people expected to be private. And in 2019, the FTC punished Facebook with a 5-billion-dollar fine.*
- Whether that has any impact is up for debate. Facebook's stock price actually rose on the day the fine was announced. The agreement also did nothing to change the core model of how Facebook monetizes users' data, leading many to believe that little will change as a result.

* This was by far the largest fine the FTC had ever given. But even such a substantial fine didn't likely impact Facebook's operations, as the company's net profit was more than 22 billion dollars in 2018.

- Facebook is not the only company to exploit user data in questionable and potentially illegal ways; it is just a large and convenient example.

This sort of behavior is common and suggests that there may be a need for different legal approaches to govern privacy.

Europe

- Europe has always had a more user-centric legal approach to privacy and personal information than the United States.
- There were a number of right-to-be-forgotten lawsuits in Europe against Google, where citizens claimed a right to have information about them removed from search results. One of the most prominent cases was filed by a man in Spain. In 1998, his home had been foreclosed on and auctioned off. By 2014, when you searched for the man's name, the auction for the foreclosure was the first result that appeared. Even though so much time had passed that the foreclosure would no longer be on a credit report, anyone who wanted to look for the man online would find evidence of his past financial issues.
- The European courts ruled in favor of the man, and Google had to comply not just for him, but for any citizen who asked to have personal information removed from search results. However, this was not universal. There were exceptions where Google could keep the results up.
- And interestingly, most requests were subject to this exception. Some of the most popular requests for information to be deleted came from public officials who wanted to have stories published in newspapers about their misdeeds taken out of the results. Because this constituted something that was of public interest, it was not covered by the right-to-be-forgotten lawsuit, and the pages were allowed to stay up.

❑ Older European privacy laws manifested for Google in interesting ways. For example, when Google announced that it was going to bring its Street View technology to Germany, there was mass outrage. While 99 percent of American streets have been mapped, allowing anyone to drop in and see a panoramic view of what that point looks like, privacy-conscious Germans were opposed to the move. In an attempt to appease the country, Google started

blurring out people, cars, and even people's homes. Anyone can request that Google blur their address. As a result, there is relatively minimal Street View coverage in Germany specifically and Europe more broadly.

❑ The reason that European citizens are able to accomplish these things is because personal privacy is at the heart of European law. In Europe, people own the data about themselves.

Even if you are in the US, you have likely seen some impacts of the GDPR. Cookie alerts at the bottom of web pages are probably the most visible. This is because the law applies to European citizens, not just European companies, so if an EU citizen views a US-based website, the US company must still comply with the law.

- Before 2018, directives from the European Union were implemented in the legal frameworks of each member country in slightly different ways. That all changed in the spring of 2018 with the implementation of the General Data Protection Regulation (GDPR)—a European-wide law that enhanced rights of European citizens and anyone who is in Europe to control their own data. Interestingly, it is not a regulation of companies that tells them what they can and cannot do, but rather a regulation that gives rights to citizens that companies have to respect.
- Again, at the heart of this is a principle that states that people own the data about themselves and have the right to control it. Specifically, they have the right to know all the data that a company has about them, to request that it be deleted, and to move that data around.
- In fact, when the GDPR came into effect, Facebook started offering the option to download personal data to everyone with a Facebook account. You can go to Facebook—or, in fact, to most social media sites—and request a download of your data. This includes all of your posts and photos and comments, but it also includes any background information that was collected about you. When the GDPR came into effect and people started downloading this information, there was quite an uproar. People were surprised by just how extensive the data collection was.
- Under the GDPR, you also have the right to request that data about you be deleted. There are some exceptions to this right. For example, your bank may not be able to immediately delete all the information it has about you because it also has a legal obligation to keep financial records for seven years for auditing purposes. However, the heart of the law says that you can go to any company and as long as there is no need for that company to keep your data, it has to delete it.

- There are also protections in place to keep companies from making up fake needs just so they can keep data about you. If companies break these laws, there are significant penalties that the European Union can levy. Citizens also have the right to file lawsuits against the companies, including class action lawsuits.

China

- China is a different place economically, culturally, and legally from the US and Europe, with different traditions and values—and privacy laws.
- Whereas one of the types of personal data collected about people in the United States is the credit score,** China does not have a credit score, and credit is used differently there. Many people do not have credit cards, and loan debt is tracked and enforced differently. That makes a US-like credit-scoring system a poor fit.
- China needed a way to assess the trustworthiness of citizens, especially when it came to loans, but potentially for other things. China's solution was the development of the social credit score. Like the US credit score, China's social credit score is a number assigned to each person. It reflects how much a person can be relied on, but not just to repay loans. While that is how it initially started, it has evolved to essentially be a measure of social standing—for responsibility or conformity, depending on who you ask. The system is still being developed and deployed, but parts of it are already in use.

** A few private agencies collect information on how timely US citizens are in repaying their debts, and that is turned into a score that is used to assess whether they are lent money, given credit cards, or qualify for a mortgage.

- It began as a way to try to force people to repay their debts. Bankruptcy doesn't exist in China the same way it does in the United States, but the penalty for not repaying debt was relatively minor. As a result, many people simply walked away from money that they owed. With the introduction of the social credit score, people who did not repay their debts and who were taken to court could be given a low score.
- If your score gets very low, such as if you suffer a large business loss and can't repay your debts, you will be put on the lowest tier of the social credit score system, called the blacklist. People who are on the blacklist have a difficult time borrowing money and opening bank accounts, which may be expected, but they are also denied all kinds of other rights, which would be very surprising to people outside of China.
- For example, they are not allowed to board most types of fast transportation, including airplanes and fast trains. There are also large digital billboards set up around cities that cycle through photographs of people who are on the blacklist and display their name and ID number to warn people about interacting with these "untrustworthy individuals." These pictures and warnings are also sometimes shown before movies in theaters.
- People who are on the blacklist can also experience degradations in service, such as having fewer channels available on cable television or having their internet service slowed down.
- How does the government know about social transgressions?*** There are neighborhood watchers who keep track of people's behavior and report it to a central system.

*** Social transgressions that can lower your social credit score include not stopping for a pedestrian who is trying to cross the street, not picking up after your dog, playing too many video games, and buying products the government thinks you shouldn't.

Also, China has a broadly deployed surveillance system, with cameras in many public places. China also has a very advanced facial recognition system that is able to identify people as they move through public places.

- The government has also partnered with many large businesses to exchange data. Alibaba, a huge corporation that is sort of like the Amazon of China, plus China's largest ride-sharing service and their largest dating site have partnered with the government on the social credit score. It is unclear exactly how information is flowing between these

companies and the government, but experts believe that the platforms are sharing data about users' habits and that the companies may be receiving scores in return.

- One thing that people can be punished for in the social credit score system is for spreading misinformation online. And the government decides what counts as misinformation, so criticizing the government online—something that already is highly monitored and punished—can be tracked and used to blacklist anyone who speaks out against the regime.

So What Can You Do?

- This is a space where awareness is especially important. Each country and state has its own privacy laws. Being aware of the laws that govern others provides insight both into what rights you could have and how bad things could get. If other citizens have more protections

over their personal data, why don't you? Are your rights being weakened in order to help companies make more money? Or are the protective laws having unintended results that are harmful? What about places where there are fewer protections? Is that a world you want to live in?

- Societies are different, so the manifestations may vary between countries, but with a lack of legal protection, similar issues can arise across cultures. Being aware of what happens in a society when different legal frameworks are in place helps you see what kind of protections matter to you and what kind of world you may face with or without them.

Assignment

Discover what privacy laws govern you. This lecture addressed national legislation, but what about your state, province, or city? Check out your local privacy laws. These may be general privacy regulations or rules about things like surveillance, drones flying over private property, or cable companies collecting data about your viewing habits. Do you have some protections that surprise you? Were you expecting more protections than you actually have? This is a good chance to brush up on your rights and to think about whether there are any issues where you may want to lobby your representatives.

Resources

Garcia, Smith, and Wang, "Episode 871: Blacklisted in China," <https://www.npr.org/sections/money/2018/10/26/661163105/episode-871-blacklisted-in-china>.

Gottlieb, *The GDPR Guy*, <https://podcasts.apple.com/us/podcast/the-gdpr-guy/id1225996374>.

Kharpal, "Apple vs FBI," <https://www.cnbc.com/2016/03/29/apple-vs-fbi-all-you-need-to-know.html>.



NAVIGATING THE FUTURE OF PERSONAL DATA

LECTURE 12

What is the future of personal data and privacy? The answer is hard to come by, not only because there are so many paths, but because technology drives this and is evolving quickly. Nevertheless, this question might be answered in a variety of ways in the coming decades.



DNA as Personal Data

- DNA represents an interesting crossroads of technological, legal, and ethical issues with personal data. DNA is now in the toolbox of every law enforcement organization to catch criminals. But in the 30 years since DNA started being used in the courtroom, science has come a long way in being able to identify and link people with their genetic profiles. And this serves as an example of how technological advances can undo privacy guarantees that were made in the past and create new challenges going forward.
- Familial DNA can be used in criminal cases.* If detectives have an unknown murderer or rapist, they could identify the person using the DNA of the individual's family members.
- In fact, this strategy helped catch an infamous serial rapist and killer who had eluded police for decades.
- The Golden State Killer** committed dozens of rapes in the Bay Area before escalating to home invasions and murder in the Los Angeles area in the 1970s and 1980s. The police had DNA but had been unable to match it to a known person in all that time. Eventually, they turned to familial DNA—using public, open databases where people voluntarily upload their DNA profiles to try to find relatives. They were searching for relatives of a murderer.
- Law enforcement was able to eventually narrow it down to one man who they thought was the killer they were looking for.

* Essentially, detectives take the profile of a person and look for whatever distant relatives they can find. Working with other public records and genealogical data brings them increasingly closer to their near relatives.


** The Golden State Killer—also known as the East Area Rapist and the Original Night Stalker—was profiled most prominently in Michelle McNamara's book *I'll Be Gone in the Dark*.

Of course, they couldn't rely just on these familial records for an arrest, but once those records gave them a suspect, they were able to surreptitiously collect a DNA sample from the door handle of the suspect's car, and it matched. The Golden State Killer was caught. It was 72-year-old Joseph DeAngelo, and he was arrested in 2018.

- The Golden State Killer was just the first prominent example in what has become a string of cold cases that have been solved with this technology. Catching killers and rapists is good—but there are a number of questions that arise about this use of personal data. In this case, people had voluntarily uploaded their profiles into public databases, where one could reasonably presume law enforcement would also have access to it. They may be surprised to know that their profiles could be used to catch their distant relatives for committing crimes.
 - When using this information leads to catching murderers and rapists, it seems we're firmly on the good side.
- But what if it starts being used for more petty crimes, such as a drug crime where DNA is left behind? What if it becomes so inexpensive that it's eventually used to catch people for committing relatively minor crimes, such as littering?
- And there are other applications where the territory becomes more troubling. For example, consider the case of anonymous sperm donation or egg donation. Most of the time, when men and women choose to anonymously donate in this way, they sign legal contracts that protect their identity. They surrender parental rights, and the contracts keep their identities private and unknown to the families that receive donated sperm and eggs.
 - If donors knew they could be identified, it's likely that many would refuse to participate. Some simply want to keep their donations to themselves. So a reasonable precaution donors might take is to keep their DNA records out of public databases. They may even choose to never get a DNA test at all.

- However, with even basic ancestral DNA searches, if that donor has a brother, sister, or parents who upload their DNA, the child that resulted from the donation would find that immediate relative quite quickly in his or her search. Then, if the child is working in a system that allows contact with genetic matches, he or she could reach out to the donor's family. If the donor never told anyone about his or her donation, the child has revealed a deeply personal and private piece of reproductive health information to family members. This is especially troublesome if the donation violates the family's ethical or religious norms.
- This kind of revelation is severe enough that it could destroy family relationships for some people, and the donor has a right to keep that information private. Furthermore, a donor likely does not want a relationship with the child that he or she anonymously donated to produce. Yet a donor's family could decide to pursue a relationship after finding a DNA match. The donor's right to control the outcome of his or her donation is taken away.
- Sperm banks and reproductive health facilities are now considering how to discuss with potential donors the way that their anonymity will be preserved. They may simply be unable to guarantee anonymity in the current world of DNA testing. But for people who donated in the 1990s and 2000s, who were assured that their identity would be kept private, familial DNA search is now potentially taking that away, and not for any real social good, but potentially for the whims and curiosities of other people.
- And of course, the problems go deeper than this. There is only a small amount of protection with respect to DNA profiles at this point.***

*** President George W. Bush signed legislation that prohibited health insurance companies from discriminating based on DNA, but it's a very narrow law.



It is worth considering whether you want to have your DNA tested by a company that offers this service. If you do want it for your own genetic insights or if you already have had it tested, you can strictly control the privacy settings on your DNA in the system. You can also have your DNA profile deleted after you have obtained the information you want from it.

- ❑ And there's a long line of discrimination based on medical misunderstandings by laypeople. For example, Ryan White was barred from attending public school in the 1980s because he was HIV positive.
- ❑ If genetic testing becomes cheap and easy, there are no current laws that prevent employers, schools, and other organizations from discriminating against people based solely on their genetic profiles. So if you are genetically predisposed toward heart disease, even if you are taking all the behavioral steps that help prevent it, you could still be barred from getting a job based on discrimination against that factor.
- ❑ The lack of scientific sophistication among the general public, who is not trained in interpreting and understanding genetic testing, means that the opportunities for unfairness are rampant, and we're likely to continue to see this sort of discrimination based on DNA.

- ❑ Genetic privacy is a complex topic, and it is unlikely that a single law could be put in place that protects people from unfairness, discrimination, and having their privacy compromised. We want to

allow for reasonable law enforcement and health-care use of DNA but protect people as we start moving toward more regulation in this space. Just how to do that is uncertain.

The Future of Privacy Regulations

- ❑ DNA is only one example of many technologies that are evolving to provide more insight and invasion into our lives. Artificial intelligence, data integration, and massive data collection all promise to lead toward new tech that can uncover identities, attributes, and connections that we never expected. As a result, we likely need to think about fundamental privacy rights that we want to establish rather than piecing together domain-specific regulations.
- ❑ US federal regulations are still largely up in the air, but there are interesting developments happening on the state level, especially in California, that may offer some insight.

- ❑ The California Consumer Privacy Act is a state law that some people refer to as “GDPR light.” Set to go into effect in January 2020, it gives many similar rights to citizens of California that Europeans have. It governs large businesses and businesses that make most of their money by sharing or processing personal data.
- ❑ The law offers a number of protections. It requires that companies be transparent about what data is collected and how they use it. Citizens have a right to control the data about themselves, and they have the right to see the data that companies hold. They have the right to request that it be deleted.

- While this is very beneficial for citizens of California, it raises the prospect of a GDPR-like federal law in the United States. That's because California is likely not going to be alone in passing a consumer privacy act. Many other states have their own privacy laws, and several are considering bills that will grant similar protections within their borders.
- Having a bunch of different state laws that regulate consumer privacy, especially when those regulations are not the same across states, makes it very difficult for a company working with personal data to operate in the United States. You potentially have to handle it differently and offer different features across 50 states. Depending on how these laws are written, you may even have to offer different protections if people are simply visiting a state versus living there.
- This scenario makes it more likely that we'll see a federal law come into place soon that offers similar protections. This would allow companies to operate under a single US privacy law as opposed to operating differently in each state. If this happens, the US will be following in Europe's lead.
- There are already precedents for federal laws to protect privacy in the US, such as the Children's Online Privacy Protection Act and the Health Insurance Portability and Accountability Act. These are federal laws that allow for consistent implementation of privacy policies around the country.

The Legislative Future of Personal Data

- The legislative future is not just limited to privacy protection laws. The US needs a robust set of laws that covers many different aspects of the personal data problem.

- ❑ Cybersecurity is a real issue that connects with data privacy. To prevent people from hacking deeply sensitive information, we need a strong defense against their attacks. That requires comprehensive cybersecurity and strong incentives for companies to follow best practices and the latest guidelines.
- ❑ Right now, the penalties for poor security practices are relatively weak. Even when companies are gathering tremendously sensitive information about people, the penalties for weak security tend to be small enough that their business is not disrupted.
- ❑ The US needs better cybersecurity laws that create very harsh penalties for companies that do not protect personal data. The US also needs to understand and discourage massive data collection without purpose.
- ❑ There is not going to be any omnibus bill that addresses all of the issues surrounding personal data. Instead, we are going to have to look at the different parts of this very complicated problem and come up with steps that improve each.

One interesting proposal that has been floated in the US Senate is to require publicly traded companies to disclose the monetary value and liability associated with the personal data they hold about people. This would serve as a financial deterrent for saving unnecessary personal information.

- The focus on corporations also carries into the discussion on privacy and surveillance. We have considered the rights of consumers and what legislation may do to protect them from companies who have their data. But what about the relationship between people and companies when the people are employees?
- Companies have very few limitations when it comes to monitoring their employees. We likely expect that companies can monitor our work emails, even though we hope they are not reading them regularly just to monitor us. But monitoring technology often extends out of the office and out of the bounds of work-related activities.
- We have become so used to surveillance in workplace situations that we may need

to step back and ask why an employer has any legitimate right to track its employees. Companies may require tracking with a variety of apps and devices for insurance coverage or discounts, and they end up with a lot of very personal data for no legitimate reason.

- A lot of the discussion around privacy rights has centered on companies and their users' data. But there should also be a serious and critical analysis of companies monitoring employees and legislation that enshrines protections for workers as well. Of course, monitoring within the workplace may be important and legitimate and should be allowed. But as employers cross into monitoring the private lives of workers, the space shifts from one of desired productivity to desired power.

So What Can You Do?

- If you are interested in learning more about privacy and security legislation across domains, there are a few steps to take. The first is to talk to

your representative, on both the state and federal levels. Ask your representatives what bills they support to protect privacy and security.

- However, it helps to have a background in what bills are being considered. For that, you might look to privacy advocacy groups such as the Electronic Frontier Foundation (EFF), which lobbies for net neutrality, better security, and better privacy laws on the state and federal levels. The EFF's website lists bills that are under consideration, where they stand, who supports them, and which of your legislators you can contact to try to swing them to support those bills. It's an excellent resource for understanding the future legal landscape on these kinds of issues.
- The technology that can collect, analyze, and derive insights from our data is growing fast. As a society, we have not figured out how to apply our ethics, values, and protections in this domain. To address this legally, we need to think about the fundamental rights people have over their data. There is change happening here as well, though the future is uncertain. But trends suggest that we may see more protections coming.

Assignment

Check to see what's going on with privacy-related legislation. The Electronic Frontier Foundation is a nonpartisan, nonprofit group that advocates for online civil liberties. The EFF is especially focused on privacy and security and how to balance these against legitimate concerns, such as law enforcement uses. You do not need to agree with all their legislative positions, but they do a good job of keeping track of the major issues that are under consideration.

Visit their website, click on Issues, and then on Privacy. Explore a topic on this page that interests you, such as medical privacy or cybersecurity legislation. This is a great opportunity to familiarize yourself with the privacy concerns related to a wide range of issues and to examine legislation that's under consideration.

Resources

Electronic Frontier Foundation, <http://eff.org/>.

Fox News, "Boy Banned from School for Carrying Cystic Fibrosis Gene," <https://www.foxnews.com/health/boy-banned-from-school-for-carrying-cystic-fibrosis-gene>.

Vincent, "Woman Fired after Disabling Work App That Tracked Her Movements 24/7," <https://www.theverge.com/2015/5/13/8597081/worker-gps-fired-myrna-arias-xora>.

RESOURCES

Allen, Marshall. "With Connected CPAP Machines, Insurers Make 'You Snooze, You Lose' Literally True." *Houston Chronicle*, November 21, 2018. <https://www.houstonchronicle.com/business/article/With-connected-CPAP-machines-insurers-make-you-13411742.php>.

Companies secretly monitor your CPAP use and send the information directly to your insurance, bypassing the doctor. It may mean your coverage is denied.

Bhat, Anjali. "Google Buzz: A Recap of the Controversy and the Current Legal Issues." *The Columbia Science and Technology Law Review*, February 19, 2010. <http://stlr.org/2010/02/19/google-buzz-a-recap-of-the-controversy-and-the-current-legal-issues/>. A look at the scandal that may have permanently doomed Google in the social networking space.

Bitcoin. <https://bitcoin.org/en/>. A cryptocurrency used for many things, including being the favored currency of the dark web.

Centers for Disease Control and Prevention. "U.S. Public Health Service Syphilis Study at Tuskegee." <https://www.cdc.gov/tuskegee/timeline.htm>. A history of the terrible Tuskegee experiments.

Chang, Alvin. "The Facebook and Cambridge Analytica Scandal, Explained with a Simple Diagram." *Vox*, May 2, 2018. <https://www.vox.com/policy-and-politics/2018/3/23/17151916/facebook-cambridge-analytica-trump-diagram>. A good primer on a complex scandal.

Dastin, Jeffrey. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women." *Business News*, October 9, 2018. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>. Using algorithms to make decisions where humans have traditionally been biased is a bad idea. This article offers an example of that at play.

Resources

Disconnect. <https://disconnect.me/>. Privacy Pro, offered by Disconnect, is a tool for identifying and blocking trackers.

Duhigg, Charles. "How Companies Learn Your Secrets." *The New York Times Magazine*, February 16, 2012. <https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>. Article about Target predicting women's pregnancies.

Electronic Frontier Foundation. <http://eff.org/>. A general, nonpartisan resource for information about legislation related to privacy and security.

Facebook. "Accessing & Downloading Your Information." https://www.facebook.com/help/1701730696756992?helpref=hc_global_nav. Facebook's guide to how you can download your data.

Fowler, Geoffrey A. "It's the Middle of the Night. Do You Know Who Your iPhone Is Talking To?" *The Washington Post*, May 28, 2019. <https://www.washingtonpost.com/technology/2019/05/28/its-middle-night-do-you-know-who-your-iphone-is-talking/>. Article about trackers on your phone.

Fox News. "Boy Banned from School for Carrying Cystic Fibrosis Gene." October 18, 2012. <https://www.foxnews.com/health/boy-banned-from-school-for-carrying-cystic-fibrosis-gene>. How a misunderstanding of personal data leads to bad results.

Garcia, Cardiff, Stacey Vanek Smith, and Echo Wang. "Episode 871: Blacklisted in China." *Planet Money*, October 26, 2018. <https://www.npr.org/sections/money/2018/10/26/661163105/episode-871-blacklisted-in-china>. A podcast taking a personal look at the impact of China's social credit score.

Ghostery. <https://www.ghostery.com/>. A tool to block trackers.

Golbeck, Jennifer. "How to Keep Your Web Browsing Private." *Psychology Today*, March 29, 2017. <https://www.psychologytoday.com/us/blog/your-online-secrets/201703/how-keep-your-web-browsing-private>. A step-by-step guide on setting up a VPN and alternative DNS server.

———. "How to Set Up a Virtual Private Network." *Slate*, February 3, 2017. <https://slate.com/technology/2017/02/how-to-set-up-a-virtual-private-network.html>. A walk-through on VPNs.

———. "Smart People Prefer Curly Fries." *Slate*, October 7, 2014. <https://slate.com/technology/2014/10/youarewhatyoulike-find-out-what-algorithms-can-tell-about-you-based-on-your-facebook-account.html>. Article about predicting personal attributes from Facebook likes.

Goldstein, Jessica M. "Meet the Woman Who Did Everything in Her Power to Hide Her Pregnancy from Big Data." *ThinkProgress*, April 29, 2014. <https://thinkprogress.org/meet-the-woman-who-did-everything-in-her-power-to-hide-her-pregnancy-from-big-data-80070cf6edd2/>. Shows just how hard it can be to keep your information private.

Gottlieb, Carl. *The GDPR Guy*. <https://podcasts.apple.com/us/podcast/the-gdpr-guy/id1225996374>. A podcast series that takes a deep look at the EU General Data Protection Regulation (GDPR) and what it means for businesses.

Haskins, Caroline. "Amazon Sent 1,700 Alexa Recordings to the Wrong Person." *VICE*, December 20, 2018. https://www.vice.com/en_us/article/pa54g8/amazon-sent-1700-alexa-recordings-to-the-wrong-person. Amazon collects a lot of data from its smart devices and doesn't seem to take great care of it.

Jackson, Amanda. "Facial-Recognition Technology Flagged 26 California Lawmakers as Criminals. This Bill to Ban the Tech Is Headed to the Senate." *CNN Business*, August 15, 2019. <https://www.cnn.com/2019/08/13/tech/facial-recognition-body-cameras-bill-california-trnd/index.html>. What happened when facial recognition was used on legislators.

Kharpal, Arjun. "Apple vs FBI: All You Need to Know." *CNBC*, March 29, 2016. <https://www.cnbc.com/2016/03/29/apple-vs-fbi-all-you-need-to-know.html>. A primer on all the issues in Apple's battle with the FBI over creating a security hole in iPhones.

Lord, Nate. "A Timeline of the Ashley Madison Hack." *Digital Guardian*, July 27, 2017. <https://digitalguardian.com/blog/timeline-ashley-madison-hack>. All the dirty details of a major hacking scandal.

Madrigal, Alexis C. "The Mysterious Printer Code That Could Have Led the FBI to Reality Winner." *The Atlantic*, June 6, 2017. <https://www.theatlantic.com/technology/archive/2017/06/the-mysterious-printer-code-that-could-have-led-the-fbi-to-reality-winner/529350/>. How hidden codes can identify printers.

Matias, Dani. "Spain's Top Soccer League Fined for Using App to Spy on Fans in Fight to Curb Piracy." *NPR*, June 12, 2019. <https://www.npr.org/2019/06/12/732157537/spains-soccer-league-fined-for-using-app-to-spy-on-fans-in-fight-to-curb-piracy>. Article about spying on Spanish World Cup fans via an app.

McLeod, Sam. "The Milgram Shock Experiment." *Simply Psychology*, February 5, 2017. <https://www.simplypsychology.org/milgram.html>. The classic obedience study that raised serious ethical concerns.

Meyer, Robinson. "Everything We Know about Facebook's Secret Mood Manipulation Experiment." *The Atlantic*, June 28, 2014. <https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/>. Facebook conducted its own psychological experiment on you without telling you or getting permission. This article explains how it worked.

Netflix Research. "Recommendations." <https://research.netflix.com/research-area/recommendations>. Netflix on their recommendations system.

O'Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books, 2016. A book about the dangers of algorithms and data, including many bias-oriented issues.

Perez, Sarah. "Some Apps Were Listening to You through the Smartphone's Mic to Track Your TV Viewing, Says Report." *TechCrunch*, January 2, 2018. <https://techcrunch.com/2018/01/02/some-apps-were-listening-to-you-through-the-smartphones-mic-says-report/>. Article about smartphone microphone usage that instructs how to check where you've consented to microphone use in apps for yourself.

Random Shopper. Tumblr. <https://randomshopper.tumblr.com/>. Dariuz Kazemi's Tumblr, where he discusses his Amazon Random Shopper bot, including the items it buys.

Rejoinder. "The Amazon Recommendations Secret to Selling More Online." <http://rejoinder.com/resources/amazon-recommendations-secret-selling-online/>. How Amazon gets huge revenue from recommending items to people.

Rubenking, Neil J. "How (and Why) to Change Your DNS Server." *PCMag*, May 17, 2019. <https://www.pcmag.com/review/364418/how-and-why-to-change-your-dns-server>. Offers information about DNS servers and lists the best ones.

Stack, Brian. "Here's How Much Your Personal Information Is Selling for on the Dark Web." *Experian*, December 6, 2017. <https://www.experian.com/blogs/ask-experian/heres-how-much-your-personal-information-is-selling-for-on-the-dark-web/>. How much data is out there about you, and what would a criminal pay to get it?

Tor. <https://www.torproject.org/>. Browser that gives you access to the dark web and some privacy protection.

TweetDelete. <https://tweetdelete.net/>. A tool to delete your older tweets automatically.

Vincent, James. "Woman Fired after Disabling Work App That Tracked Her Movements 24/7." *The Verge*, May 13, 2015. <https://www.theverge.com/2015/5/13/8597081/worker-gps-fired-myrna-arias-xora>. Should your employer be able to track you, even when you're not working? One business thought so.

What Is My IP Address. <https://whatismyipaddress.com/>. A tool that will show you the IP address of your computer and the general location of it. This information is always accessible, so it can be used by any site that you visit online.

Whittaker, Zack. "Judge Orders Amazon to Turn Over Echo Recordings in Double Murder Case." *TechCrunch*, November 14, 2018. <https://techcrunch.com/2018/11/14/amazon-echo-recordings-judge-murder-case/>. The case of the murder of Christine Sullivan and Jenna Marie Pellegrini.

IMAGE CREDITS

Page#	Credit
4	AntonioGuillem/iStock/Getty Images.
8	fizkes/iStock/Getty Images.
13	Chinnapong/iStock/Getty Images.
14	COMiCZ/iStock/Getty Images.
17	Ridofranz/iStock/Getty Images.
18	vm/E+/Getty Images.
24	Eva-Katalin/E+/Getty Images.
29	Daisy-Daisy/iStock/Getty Images.
35	fizkes/iStock/Getty Images.
37	NicoElNino/iStock/Getty Images.
44	yucelyilmaz/iStock/Getty Images.
47	RobertAx/iStock/Getty Images.
50	SDI Productions/E+/Getty Images.
55	D-Keine/E+/Getty Images.
57	izusek/iStock/Getty Images.
65	Chainarong Prasertthai/iStock/Getty Images.
69	gorodenkoff/iStock/Getty Images.
75	Mark Ramsay/Flickr/CC BY 2.0.
77	Orbon Alija/E+/Getty Images.
85	fstop123/E+/Getty Images.
95	mattjeacock/DigitalVision Vectors/Getty Images.
102	OlegAlbinsky/E+/Getty Images.
105	leolintang/iStock/Getty Images.
115	metamorworks/iStock/Getty Images.
118	anyaberkut/iStock/Getty Images.
Icons	denkcreative/DigitalVision Vectors/Getty Images.
	cnynthz/DigitalVision Vectors/Getty Images.