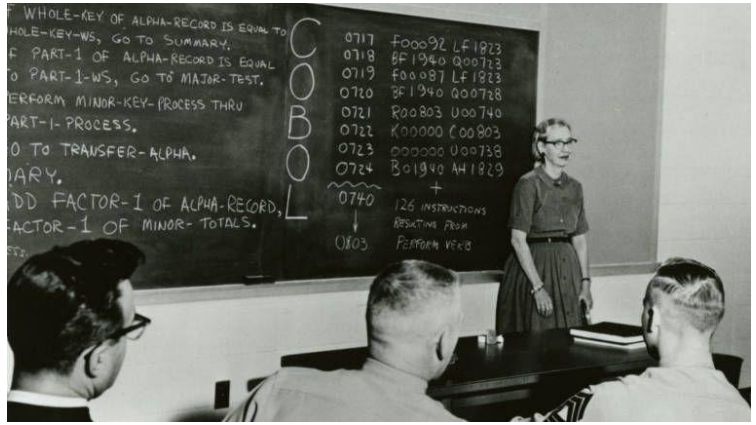
Bonus Slide

Grace Hopper and the Compiler

In 1952, Admiral Grace Hopper invented a **compiler** for the A-0 programming language that could take statements written in strictly formatted English and translate them into computer-readable code. This made it much easier to write new programs for computers. This was one of the very first compilers.



She also was part of a team that designed COBOL, one of the first plain-language programming languages.



Unit 5: CS Ethics

Ethics of Data Collection

15-110 – Friday 10/23

Announcements

- Project1 due next Monday
  - **Demo**: how to download a GenAI log

- Final Exam – **Monday 12/07 8:30am**

Learning Goals

- Recognize the current extent of **data collection** on the internet and how it impacts **data privacy**
- Understand the risks of **data aggregation**, especially related to **third party cookies** and **fingerprinting**
- **Formulate arguments** for and against regulating data collection / data privacy
- Recognize common approaches you can use to **protect your own data**

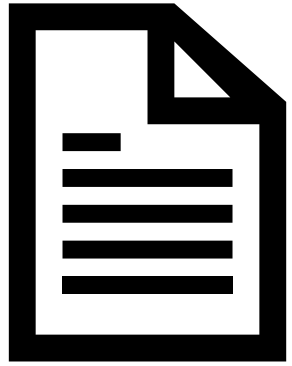
Data Collection

Data Collection Is Widespread

Every website or app that you use collects extensive **data** about you and other users.

This is stored using a **cookie**, a small piece of data that a website asks your browser to store. This data is sent back and forth between your computer and the website every time you visit it.

Some cookies used for data collection are purely functional, like tracking of site preferences or favorites. But much data is also gathered for **commercial** purposes.





Data Improves Advertising

Many websites and apps are provided to users for free. The creators still need to pay for development and server costs, and they often use **advertising** to do this.

Advertisers want to show ads to the users who are most likely to buy their product or service. How can they identify those users? **Data!**

Advertisers will require that their ads be shown to **targeted audiences** based on certain demographic/interest/behavior categories. Websites will then choose which ad to show based on the data they've collected on the current user.

Everyday Data Collection

Some data collected by websites is provided by the user voluntarily.
Other data is provided automatically by your computer.

Example: https://webkay.robinlinus.com/

You Do: Everyday Data Collection

You do: review the data big websites have about you. Does anything surprise you?

Google - https://myadcenter.google.com/controls

Amazon - https://www.amazon.com/slc/hub?useAiMode=1&ref_=ya_d_ph_ayp_1&disablePicker=1&shoppingAs=1

Instagram - https://accountscenter.instagram.com/ads/

TikTok – Profile > Menu: Settings and privacy > Ads > Manage ad topics

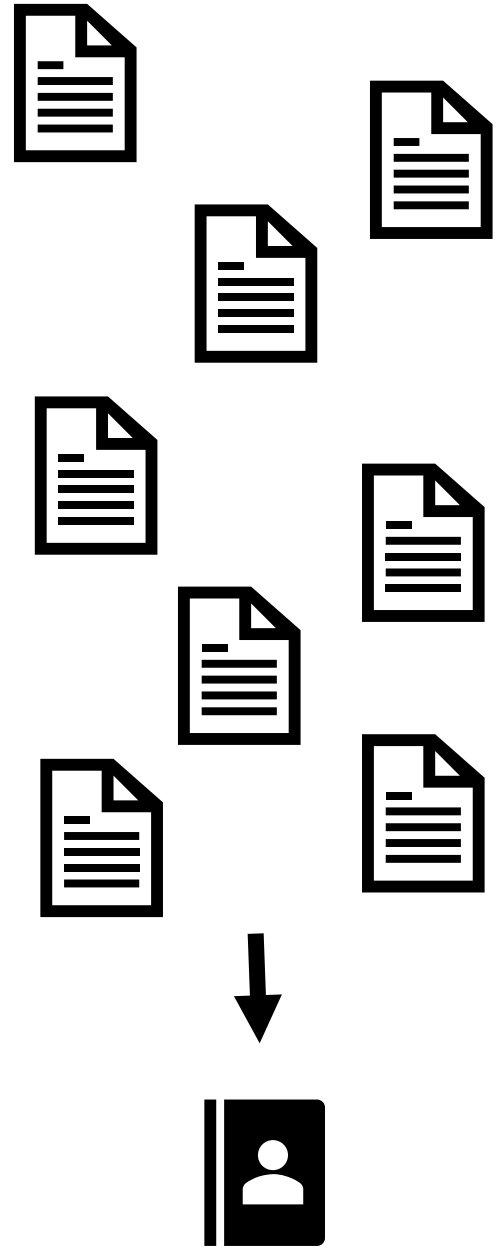
Snapchat – Settings > My Account > Ad Preferences

Data Aggregation

The risks of data collection grow greater when it expands to **data aggregation**. This happens when a company intentionally gathers and combines data from many different sources to create a consolidated database on many users.

One site usually only knows one aspect of a user's interests – for example, a music streaming site has data on your musical interests. A data aggregator typically has a much more thorough profile of a user's interests and background.

Here's one example of a data aggregator that can create a background report on a person based on just their name and city in a minute: https://www.peoplefinders.com/



Third Party Cookies Support Aggregation

One way companies collect data is through **third party cookies**.

This is a cookie requested not by the main website you're interacting with, but by a **third party** that is hosting some content on that website (often an ad). The third party will ask your browser to remember some unique ID.



That same third party will host ads on many other websites and use the same cookie across all of them. When your browser responds with the cookie originally used on a different website, the third party can **connect** the two browsing sessions under a single account.

Trackers in Real Websites

Let's see how trackers are being used in real websites you might visit:

https://themarkup.org/blacklight

Fingerprinting Support Aggregation

Even if you don't allow cookies, **fingerprinting** can be used to identify the same user across different browsing sessions.

A **device fingerprint** is a collection of information automatically provided by a browser during web requests that is unique enough to identify a user. For example, your browser's version + timezone + screen size + installed fonts might be unique when compared to the same information on any other person's computer.



See your device's fingerprint: https://coveryourtracks.eff.org/

Data Privacy

Why does it matter if companies collect or even share our data?

The principle of **data privacy** says that a person should have control over their personal data, including how companies use it.

This is an extension of the general right to privacy, which has been recognized by many institutions as a human right. For example, from the UN's Universal Declaration of Human Rights:

- No one shall be subjected to arbitrary interference with his **privacy**, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

https://www.ibm.com/think/topics/data-privacy
https://www.un.org/en/about-us/universal-declaration-of-human-rights

Case Study: Pregnancy Detection

Consider a sensitive topic: early pregnancy. Many women do not widely share that they are pregnant until they reach the second trimester (3+ months), or sometimes even later.

Large stores are able to combine volunteered information (due dates on baby registries) with collected data (common purchases) to identify purchase trends that indicate a woman might be pregnant. They can use this information to target advertisements for baby-related items to potentially-pregnant women in the future.



https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Case Study: Privacy Concerns

The big-box store Target used data to identify and send ads to expecting mothers back in 2012. However, they realized that some women reacted badly when it seemed that Target knew they were pregnant **before they told the store**.

In one anecdote, a pamphlet with baby-related items was sent to the house of a teen girl. Her father initially contacted the store manager in outrage; later, he apologized after learning his daughter actually was pregnant.

In this case, a targeted advertisement might have revealed information about the girl that she wasn't yet ready to share.

https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html

Discuss: Data and You

You do: are there any types of data that you would want to keep private?

When should there be restrictions on how companies can share data?

Regulation

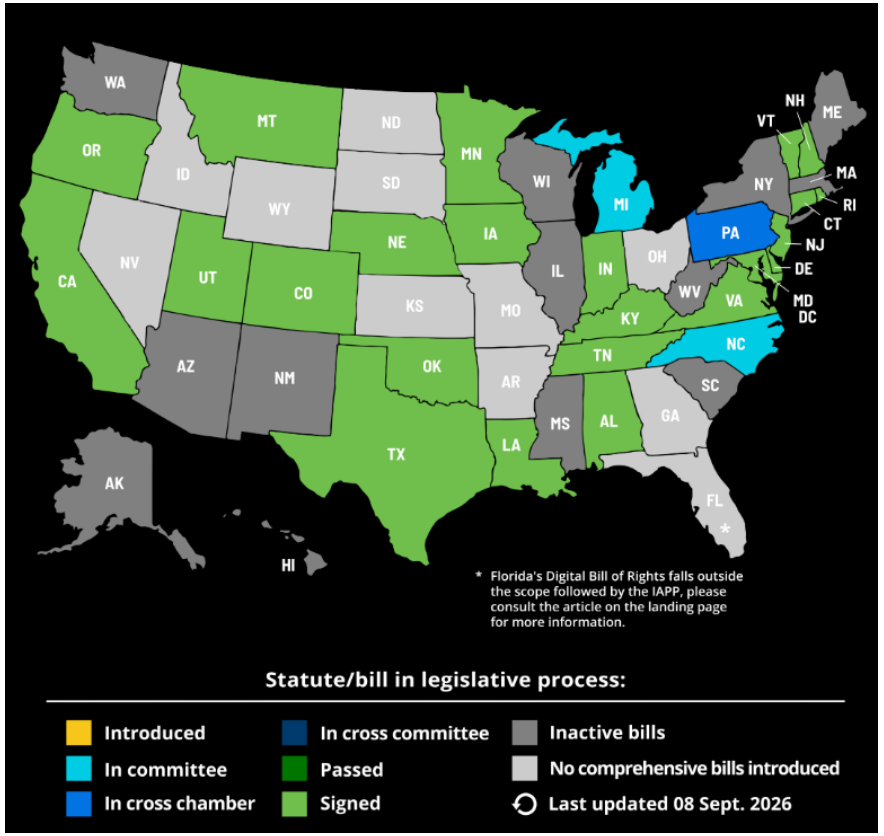
United States Regulations

The United States has many laws about privacy, but only one federal law about general data privacy.

- **COPPA** (Children's Online Privacy Protection Act) – puts restrictions on collection and use of data from children under the age of 13.

Individual states can pass their own laws, and some have.

- **CCPA** (California Consumer Privacy Act) – California residents have the right to know when data is collected, say no to their data being sold, access and correct data, and request that data be deleted.



https://iapp.org/resources/article/us-state-privacy-legislation-tracker

International Regulations

Outside the US, many countries have set their own laws over how data should be protected.

A standout example if **GDPR** (General Data Protection Regulation), which was adopted by the EU in 2016 and applies to any data collected from people in the EU.

GDPR sets principles for how data should be collected, establishes privacy rights that users have, puts limitations on when data can be processed, and generally emphasizes the importance of user consent.

Debate: Data Privacy Regulation in the US?

Debate question: should the United States pass a federal law requiring more regulation around data privacy?

Choose an answer on the scale from 1 (no regulation) to 5 (strict regulation).

https://bit.ly/110-f26-regulation-1





No regulation

Strict regulation

You Do: Background Review

Let's consider this question through the lens of the GDPR rollout. How effective has this regulation been? What challenges has it faced? Choose one of the following articles and skim through it, then discuss with a group.

- 2020 - European Commission – EU data protection rules empower citizens and are fit for the digital age
  - https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1163
  - 2024 Follow-up: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2024%3A357%3AFIN&qid=1721897017650
- 2020 - GWU Regulatory Studies Center - Unintended Consequences of GDPR
  - https://regulatorystudies.columbian.gwu.edu/unintended-consequences-gdpr
- 2021 - Internet Policy Review - Before and after GDPR: tracking in mobile apps
  - https://policyreview.info/articles/analysis/and-after-gdpr-tracking-mobile-apps
- 2024 - George Mason Law Review - A Report Card on the Impact of Europe's Privacy Regulation (GDPR) on Digital Markets
  - https://lawreview.gmu.edu/forum/a-report-card-on-the-impact-of-europes-privacy-regulation-gdpr-on-digital-markets/
- 2024 - MIT - GDPR reduced firms' data and computation use
  - https://mitsloan.mit.edu/ideas-made-to-matter/gdpr-reduced-firms-data-and-computation-use
- 2024 – EU Agency for Fundamental Rights - GDPR in practice – Experiences of data protection authorities
  - https://fra.europa.eu/en/publication/2024/gdpr-experiences-data-protection-authorities

You Do: Pros and Cons

Regulation Pros:

Regulation Pros:

Debate: Data Privacy Regulation in the US?

Debate take two: should the United States pass a federal law requiring more regulation around data privacy?

Choose an answer on the scale from 1 (no regulation) to 5 (strict regulation).

https://bit.ly/110-f26-regulation-2





No regulation

Strict regulation

Protecting Your Data

Password Policies

To keep your data safe from **malicious third parties**, keep your accounts secure.

Use long, non-dictionary-word passwords to ensure that no one can guess your password.

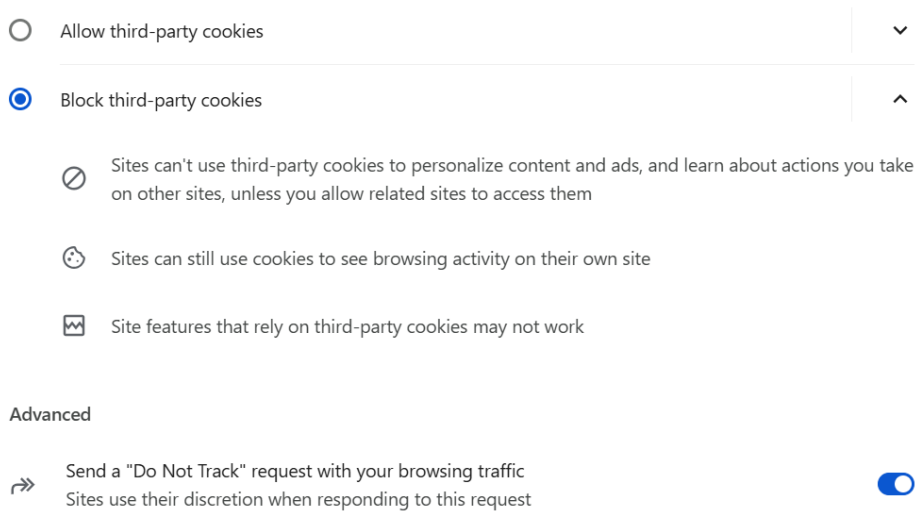
Ensure that passwords for essential services (email, bank, etc) are unique so that if another website has a data leak, your information on other sites will still be safe.

Secure Browsing

Choose browser settings that protect your data.

Many browsers let you block third party cookies and send 'Do not track' requests to websites.

When entering sensitive data, check that you're using an HTTPS (secure) connection, so that the data cannot be openly read by others.

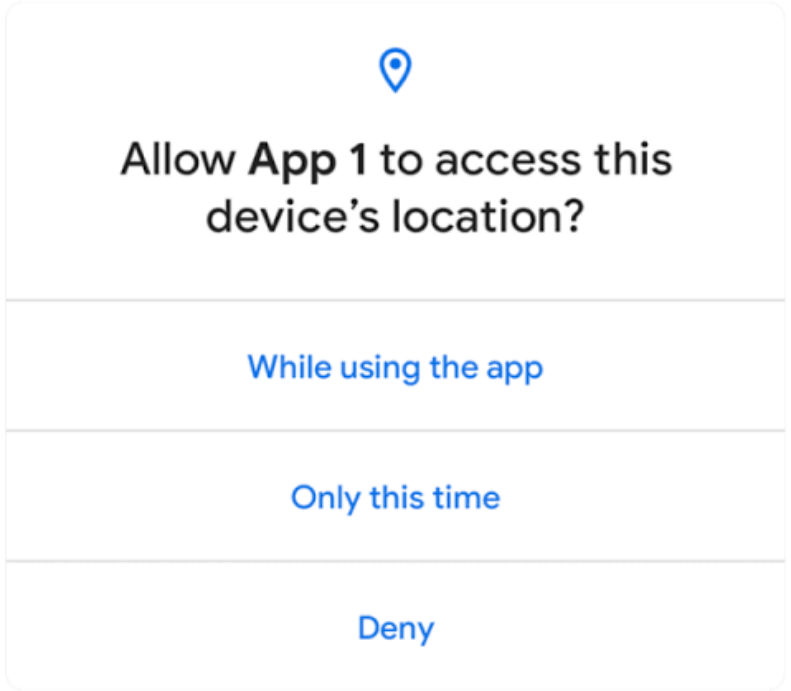


Selective Permissions

Websites and apps may request permission to access sensitive data, like your location or your camera feed.

Use discretion when assigning permissions. Does the service actually need that permission? If yes, does it need it permanently, or only for this visit?

Similarly, use discretion when deciding whether to save information (like a credit card or address) to your profile on a website.



VPNs

To further reduce the likelihood of your data being aggregated, you can use a **VPN** (Virtual Private Network) to protect your data from your ISP (Internet Service Provider), which normally can see all of your traffic on the internet.

CMU has its own VPN: https://www.cmu.edu/computing/services/endpoint/network-access/vpn/how-to/

Learning Goals

- Recognize the current extent of **data collection** on the internet and how it impacts **data privacy**
- Understand the risks of **data aggregation**, especially related to **third party cookies** and **fingerprinting**
- **Formulate arguments** for and against regulating data collection / data privacy
- Recognize common approaches you can use to **protect your own data**