

## Lect. #19: Privacy 2



[dilbert.com/strips/comic/2018-04-30/](https://dilbert.com/strips/comic/2018-04-30/)

# Agenda for Today's Lecture



## 1. Data



## 2. NPI vs PPI



## 3. Privacy Regulation in the United States



## 4. Fair Information Principles and More

# Announcements

---

- GCA-PM plan due before 11:59 pm tomorrow, 10.28.22
- Next Thursday, 11.3.22, we'll have our next speaker, Emma Clawson, Engineer at SEL

## Poll

If you have a Facebook account, have you changed your privacy settings from the default settings?

A. Yes

B. No

C. I don't know or I don't remember

D. I don't have a Facebook account

# Privacy: Nissenbaum's Privacy Model

## Contextual Integrity

- Norms of appropriateness:
  - Is a given type of personal information appropriate or inappropriate to divulge within a particular context?
- Norms of distribution:
  - Is the flow of information within or across contexts limited?

**If both norms are respected, then according to Nissenbaum, privacy has been maintained**

## Breakout Discussion (3 min)

*SuperMart installs several closed-circuit cameras in each of its aisles. These closed-circuit cameras not only allow the store to obtain aggregate statistics on how much time shoppers spend in each aisle (for sales/marketing purposes); they also allow the store to learn about the movements and preferences of individual shoppers. As a shopper walks down an aisle, a facial recognition algorithm attempts to recognize the shopper's face by matching video stills with a database of (identified) face shots that SuperMart collects when customers register for its "Club Card."*

*If SuperMart can recognize an individual shopper, that shopper's particular preferences and interests are inferred based on how long they linger in a particular place and what products they pick up and either put in their cart or back on the shelf. In addition, SuperMart uses point-of-sale receipts to collect additional information on customers' preferences. SuperMart then uses all this information to build a profile of its individual customers. The profile, in turn, is used to furnish customers with in-store targeted coupons and online targeted ads. Note that SuperMart also shares its customer information with third parties, who then use the information to furnish targeted online ads on their sites.*

- In your breakout discussion, use Nissembaum's model to analyze the scenario above to determine whether contextual integrity is maintained (use Norms of Appropriateness and Norms of Distribution)

# Privacy: Nissenbaum's Privacy Model

## SuperMart Scenario

- **Norms of appropriateness:** Is it appropriate to collect information on customer shopping behavior within SuperMart?
  - By contract theory, we can argue that SuperMart has a right to collect personal information; it's a common practice
  - *With informed consent*, SuperMart should be able to use video data to extract shopping behaviors; *if consent has not been given, then this violates norms of appropriateness*
  - Even if consent has been given, SuperMart should delete the video data after it has been used

# Privacy: Nissenbaum's Privacy Model

## SuperMart Scenario

- **Norms of distribution:** Is it appropriate to share customers' personal information, including shopping behavior, with third parties?
  - *With informed consent*, SuperMart can share personal data such as name, address, phone number, and purchases with third parties **if these third parties have privacy policies similar to that of SuperMart**
  - *With informed consent*, SuperMart can share certain data extracted from video footage (e.g., purchases considered and made) **with third parties having equivalent privacy policies**; however, sharing actual video data does not satisfy norms of distribution



# Privacy: Big Data

- Working definition of big data: analysis of large and/or complex data sets using a series of techniques
  - Three key factors: size, complexity, and analysis tools
  - Three V's: variety, velocity, and veracity
    - variety: wide range of sources involved in data analysis, e.g., social media, scientific applications, business transactions, internet search indexing, medical records, web logs
    - velocity: speed in which data generated, distributed, collected
    - veracity: quality of the data
- Raises serious privacy concerns; some think big data requires a new legal category of privacy (group privacy)

# Privacy: Data Mining

- Data mining: technique involving manipulation of information, including personal information, through analysis of implicit patterns in large data sets
- Data mining raises unique concerns for personal privacy because of what can be learned about an individual
- Both informal guidelines and formal laws exist for data protection of personal data:
  - Guidelines for electronic records (FDA)
  - Laws for medical, academic, and financial records

## Privacy: Data Mining (cont.)

- Few legal or normative protections apply to personal data manipulated by data mining where personal data is:
  - Implicit in the data
  - Nonconfidential
  - Not necessarily exchanged between databases
- Example: Credit worthiness now being determined using datamining:
  - Computer consultant had two American Express cards cancelled and the limit on his third card reduced based on where he shopped and the financial institution holding his mortgage

# Privacy: Web Mining



- Web mining: applications of data mining techniques to find patterns from the web
- Often used for online advertisements
- Facebook's Beacon initiative in 2007 used web mining to allow FB friends to share information about their online activities and purchases, but this also allowed targeted marketing; cancelled in Dec. 2007, but web mining probably still done by FB

# Privacy: NPI vs PPI

- **NPI: non-public personal information**

- Confidential, sensitive personal information
- Medical, financial, academic records
- Both normative and legal protection

- **PPI: public personal information**

- Not confidential and not considered to be sensitive
- Where you work or attend school; what kind of car you drive; what you do for fun; where you shop online; where you browse online; how much you paid for your house; what your real estate taxes are; your phone number; and more
- Need for protecting personal privacy in public


## Breakout Discussion (4 min)

- In your breakout discussion, brainstorm ideas on how your activities and data online, i.e., public personal information (PPI), might be used to violate your privacy

# Privacy: Examples of PPI

- In 2005, the Bush administration required Google and Yahoo! to submit a list of all user queries entered into their search engines during a one-week period; suppose a student, Ari, was working on a research paper on child pornography on the internet; any queries he made could have been totally misinterpreted as an interest in viewing child pornography
- A website used to exist for people to report sightings of famous people; GPS software then provided users with precise locations and times so these people could be stalked electronically or even physically

# Privacy: Examples of PPI

- Google saves every search query together with IP information on where the search occurred; it uses a deep learning algorithm for natural language processing (NLP) to evaluate all its searches 
- Alphabet owns Google, gmail, YouTube, Picasa, Fitbit (purchased last year!), Nest, and hundreds of more companies; prior to 2012, each company had a separate privacy policy, but now one privacy policy covers all its companies
- Alphabet can use all the data collected from any of its companies, representing a **GINORMOUS** amount of data from diverse sources



# Privacy: An Aside

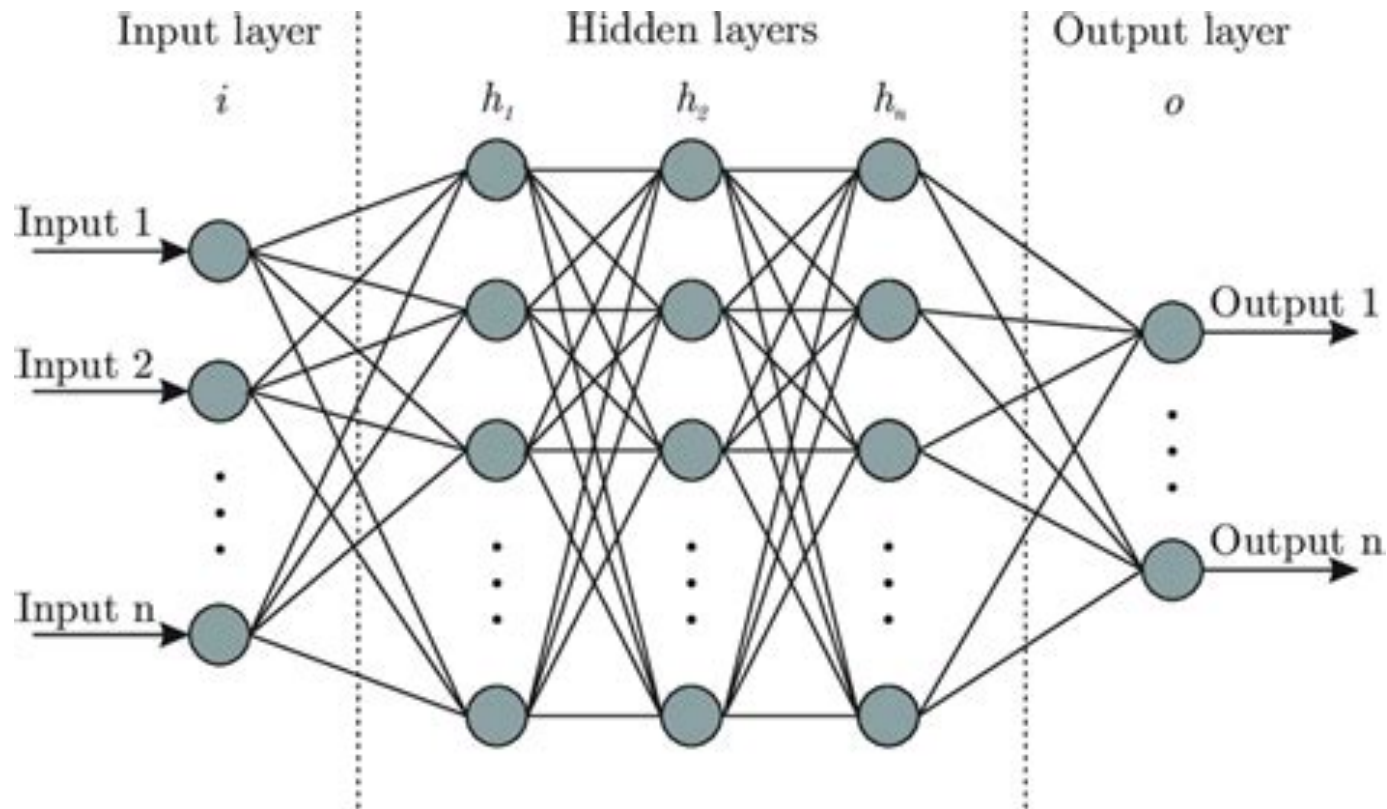
## Pokemon Go (2021)



- Internal start-up at Google headed by person in charge of Google Street View
- Now separate company called Niantic Labs
- 1+ billion downloads
  - Median age 33
  - 43% 25-44 year olds
  - 41% female
  - 71% households with kids
- Sponsored locations
  - Companies pay: “Drive foot traffic to your business through gameplay”

# Privacy: Examples of PPI

## Neural Network (Deep Learning)

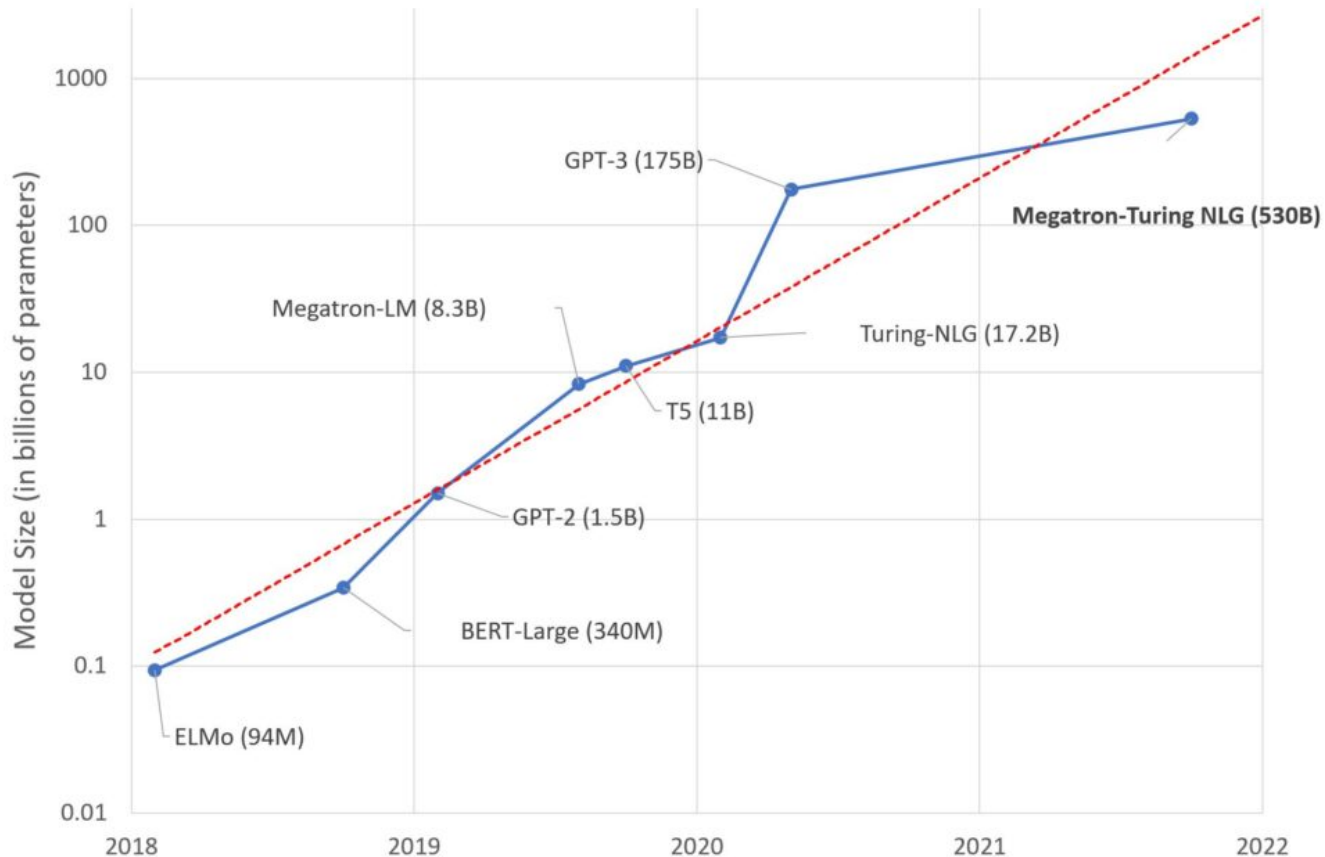


Total number of parameters = sum of weights and biases

For  $n$  input nodes,  $n$  hidden layers each with  $n$  nodes, and  $n$  output nodes, total number of parameters is  $n^2(n+1) + n(n+1)$  (if I did the calculation correctly!).

# Privacy: Examples of PPI

## Microsoft Turing Natural Language Generation (2020)



Megatron-Turing NLG: 530 billion parameters

[turing.microsoft.com](https://turing.microsoft.com)

# Privacy: National Privacy Laws in the U.S.

- **FERPA (1974):** Applies to academic records
  - Family Educational Rights and Privacy Act
- **ECPA (1986):** Applies to email
  - Electronic Communications Privacy Act
- **HIPAA (2003):** Applies to medical records
  - Health Insurance Portability and Accountability Act
- **FACTA (2003):** Applies to credit reports and more
  - Fair and Accurate Credit Transactions Act
- **GINA (2008):** Applies to personal genetic information
  - Genetic Information Nondiscrimination Act
- ***BUT NO NATIONAL PRIVACY OR PROTECTION LAWS FOR ONLINE DATA***

## Poll

Have you deleted or deactivated a social media account due to concerns over your privacy?

- A. Yes
- B. No
- C. Other

# Privacy: Fair Information Practices

## Five core principles

1. Consumers should be given notice
2. Choices should be offered and consent required
3. Consumers should be allowed to access and alter data
4. Data should be accurate and secure
5. Mechanisms for enforcement and redress are necessary

# Privacy: Principles of Informed Consent

- Informed consent means that a user is fully informed and makes decisions about their personal data
- The five principles needed for informed consent are:
  1. Disclosure: Explanation of privacy policy
  2. Competence: Ability to understand disclosure
  3. Comprehension: Understanding of privacy policy
  4. Voluntariness: Degree to which user has choice in accepting privacy policy
  5. Agreement: Agreeing with the privacy policy
- Privacy paradox: Users are concerned about online privacy, but they provide a lot of personal information

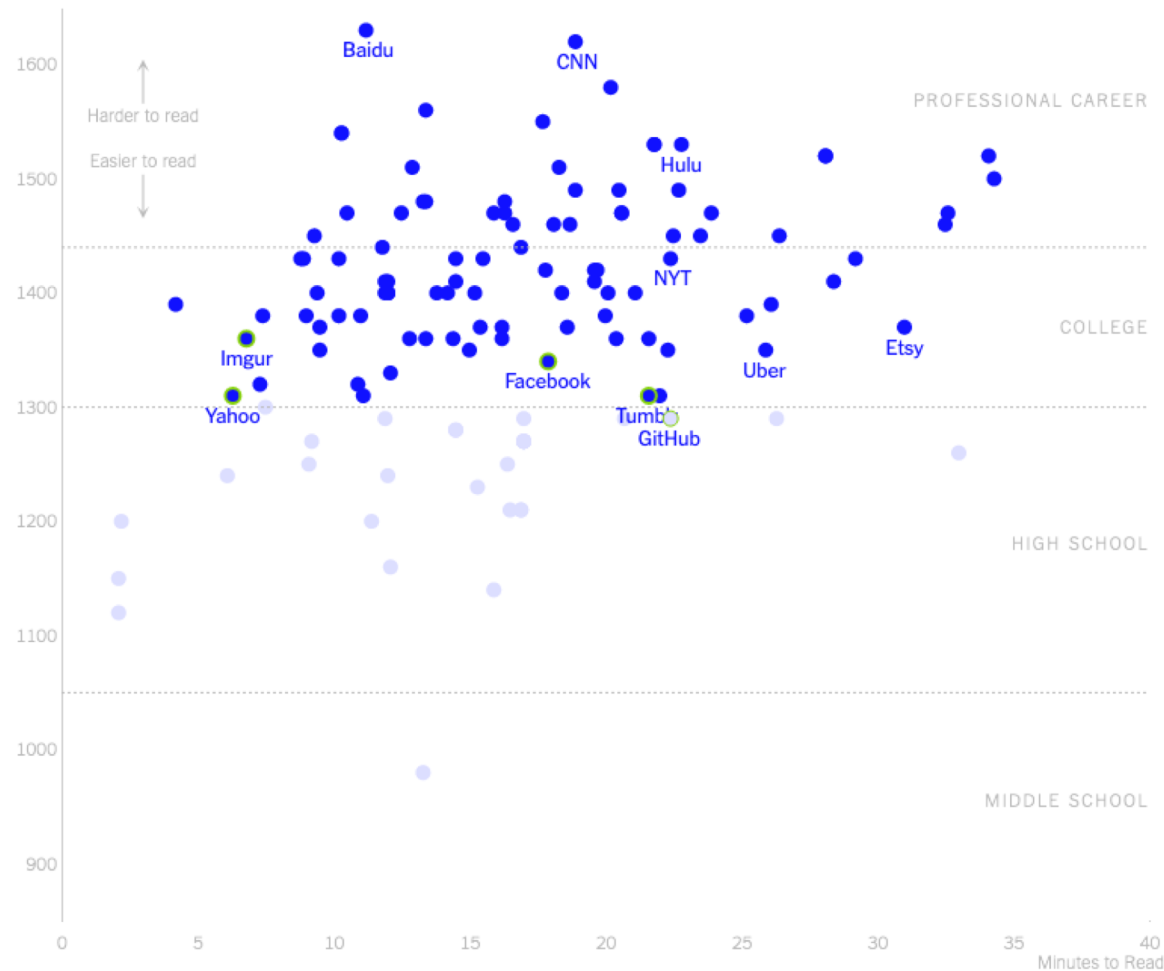
## Poll

How often do you read the privacy policies and terms of service for apps and web accounts ***completely***?

- A. Always
- B. Sometimes
- C. Rarely
- D. Never



<https://www.nytimes.com/interactive/2019/06/12/opinion/facebook-google-privacy-policies.html>



# Privacy: Privacy Policies of Major Online Companies (2017)

	Tracking cookies	Collects Unique device IDs	Scans user content	Logs user activity	Collects Location	Collects friend /contact lists	Uses localization	Uses recommendations/promotions	Uses 1st party ads	Uses 3rd party ads	Shares personal info	Shares de-identified info	Stores info in multiple countries	Indefinite retention
Google	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Apple	●	●		●	●	●	●	●	●	●		●	●	●
Microsoft	●	●	●	●	●	●	●	●	●	●		●	●	●
Facebook	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Reddit	●	●	●	●	●		●	●	●			●	●	
Twitter	●	●	●	●	●	●	●	●	●	●		●	●	
LinkedIn	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Instagram	●	●	●	●	●	●	●	●	●	●		●	●	●
WhatsApp		●		●	●	●							●	
Snapchat	●	●	●	●	●	●	●	●	●	●	●	●	●	
Amazon	●	●	●	●	●		●	●	●	●		●	●	●
Ebay	●	●	●	●	●	●	●	●	●	●	●	●	●	●
Netflix	●	●	●	●	●		●	●		●		●		●
Hulu	●	●	●	●	●	●	●	●	●	●		●		●

# Privacy: A Good Privacy Policy Disclosure Example

Reasons for sharing your personal information	Do we share the information?	Can you limit the sharing?	How can you limit the sharing?
For everyday business purposes	Yes	No	
For joint marketing with our partners	Yes	Yes	Go to <URL> or call [number]
For our affiliates' everyday business purposes	Yes	Yes	Go to <URL> or call [number]
For our nonaffiliates everyday business purposes	No	We don't share your information	