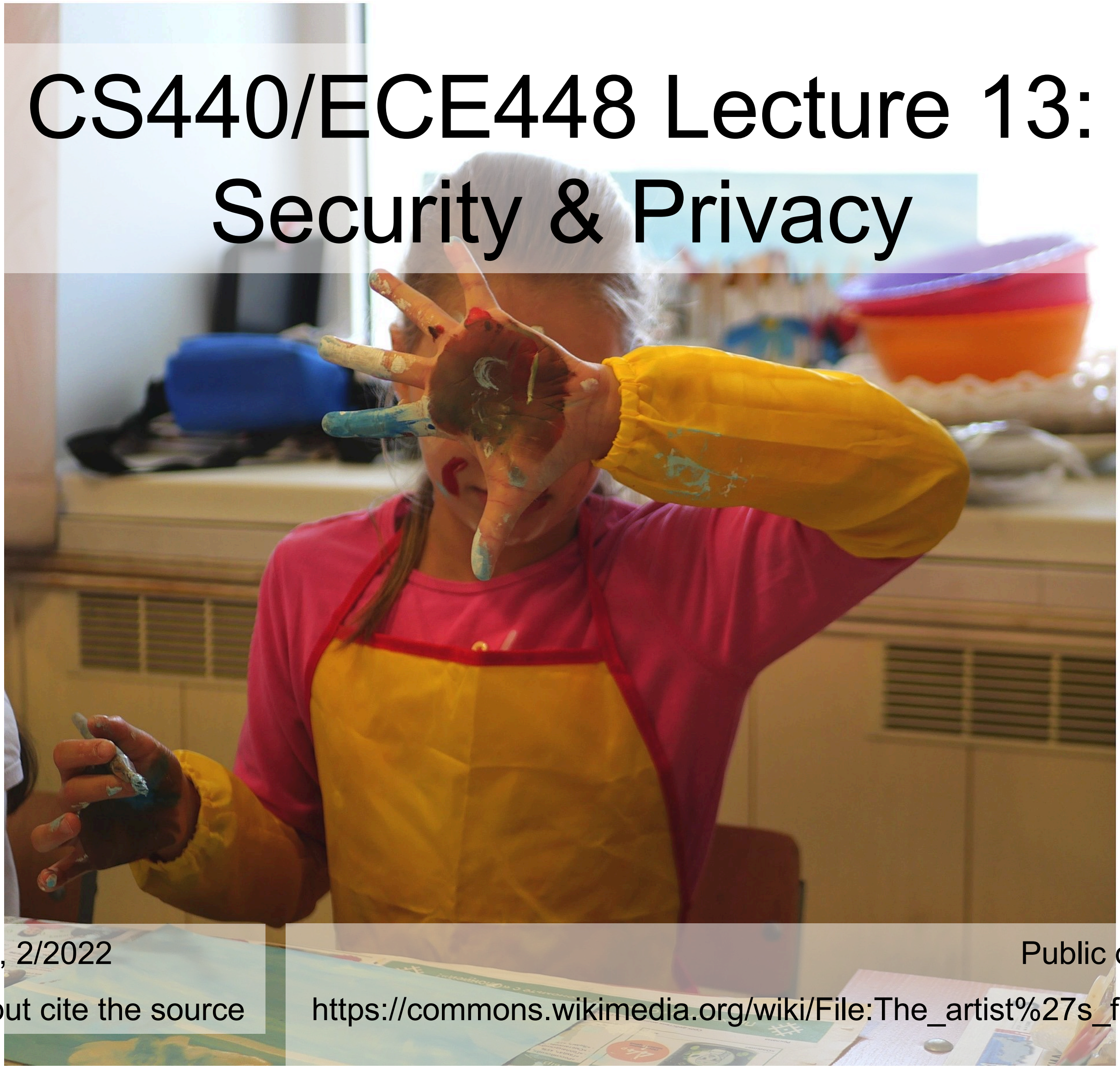
CS440/ECE448 Lecture 13: Security & Privacy



Mark Hasegawa-Johnson, 2/2022

Public domain image, AKA 2013, https://commons.wikimedia.org/wiki/File:The_artist%27s_face_behind_his_hand.jpg

Outline

- Universality and variations in the concept of privacy
- Benefits of shared and open data
- Harms caused by shared and open data
- Consent
- Technological alternatives to data sharing: differential privacy, federated learning

Is privacy a modern Western idea?

”Privacy is the ability of an individual or group to seclude themselves or information about themselves, and thereby express themselves selectively….

The concept of universal individual privacy is ***a modern concept primarily associated with Western culture***, particularly British and North American, and remained virtually unknown in some cultures until recent times.”

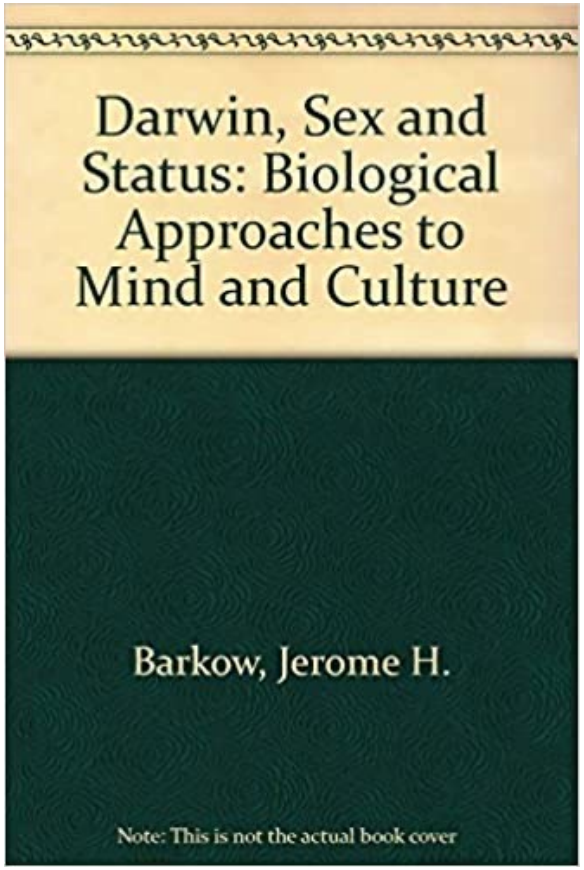
- Wikipedia

“The evidence for peoples seeking to manage the boundaries of private and public ***spans time and space, social class, and degree of technological sophistication***. Privacy – not merely hiding of data, but the selective opening and closing of the self to others -- appears to be both culturally specific and culturally universal.”

- Acquisti, Brandimarte & Hancock, *Science* 6578:270-272, 2022

Possible biological origins of privacy

- Humans may have evolved self-awareness primarily b/c it helps us better manage how others see us.
- In turn, by better managing how others see us, we improve our chances of survival.



Darwin, Sex and Status: Biological Approaches to Mind and Culture

Barkow, Jerome H.

Note: This is not the actual book cover

© University of Toronto Press, 1989

Alternatives to the modern conception of privacy

- In societies where individuals are rarely alone, other mechanisms exist to distinguish public vs. private persona.
- For example, there are spaces near Uluru (Ayers Rock, Australia) that cannot be photographed, and that can only be entered by Aṉangu members of the appropriate gender, because they are reserved for gender-specific rituals.

Helicopter view of Uluru/Ayers Rock. Hunster, 2007.
Public domain image.
https://commons.wikimedia.org/wiki/File:Uluru_(Helicopter_view)-crop.jpg

Alternatives to the modern conception of privacy

- In most pre-modern societies, high social class meant never being alone.
- Acquisti et al. (Science, 2022) note that in such situations, ”people manifested their privacy needs through stiff social interactions.”

東京藝術大学日本画研究室 現状模写

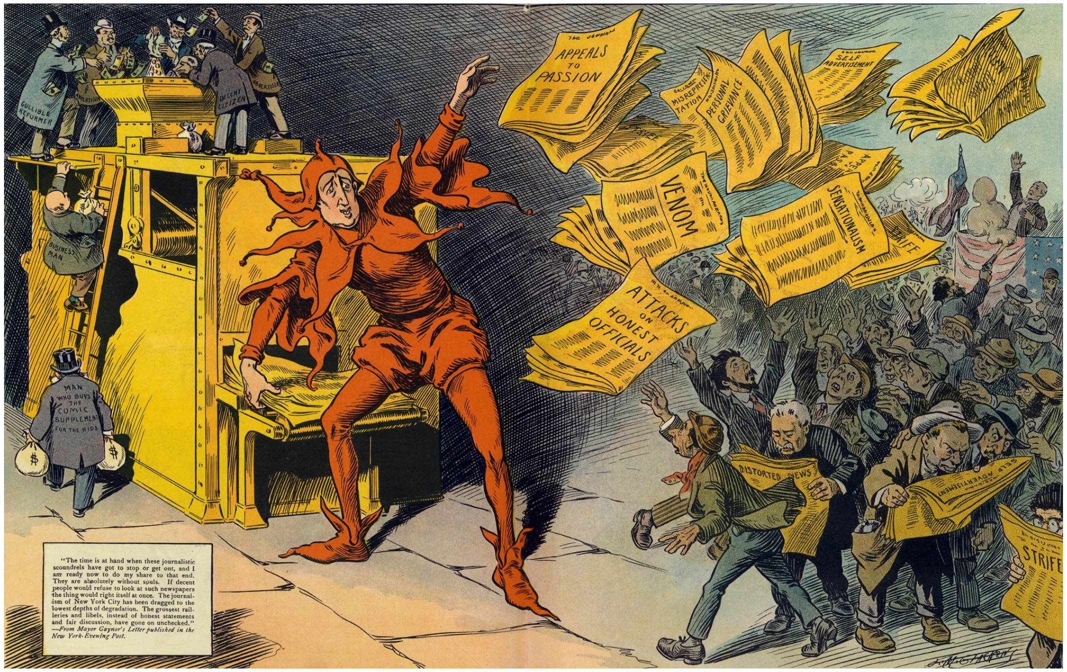
夕霧 *Yūgiri* ("Evening Mist”), 12th century scroll, Tale of Genji

Public domain image, Gotoh museum,

https://commons.wikimedia.org/wiki/File:Genji_emaki_01003_009.jpg

Origin of ”The Right to Privacy” in the U.S.

- US Constitution, Bill of Rights say nothing explicitly about privacy.
- Between 1850 and 1890, # of newspapers in U.S. grew from 100 to 900, in part b/c of sensationalist gossip about celebrities.
- Warren and Brandeis, 1890, wrote “The Right to Privacy,” an article claiming that such a right exists.



“The time is at hand when these journalistic scoundrels have got to stop or get out, and I am ready now to do my share to that end. They are absolutely without souls. If decent people would refuse to look at such newspapers the thing would right itself at once. The journalism of New York City has been dragged to the lowest depths of degradation. The grossest railleries and libels, instead of honest statements and fair discussion, have gone on unchecked.” —*From Mayor Gaynor's Letter published in the New York-Evening Post.*

The Yellow Press, by L.M. Glackens, 1910.
Public domain image,
https://commons.wikimedia.org/wiki/File:The_Yellow_Press_by_L.M._Glackens.jpg

Is privacy useful?

Solove claims it's in government's best interest to enforce a right to privacy, because:

- public/private separation permits your public persona to be simplified, e.g., to fit the attributes of a known societal role,
- simplified public persona makes economic and political transactions more efficient.

Cover of *Understanding Privacy*, © Harvard University Press, 2010

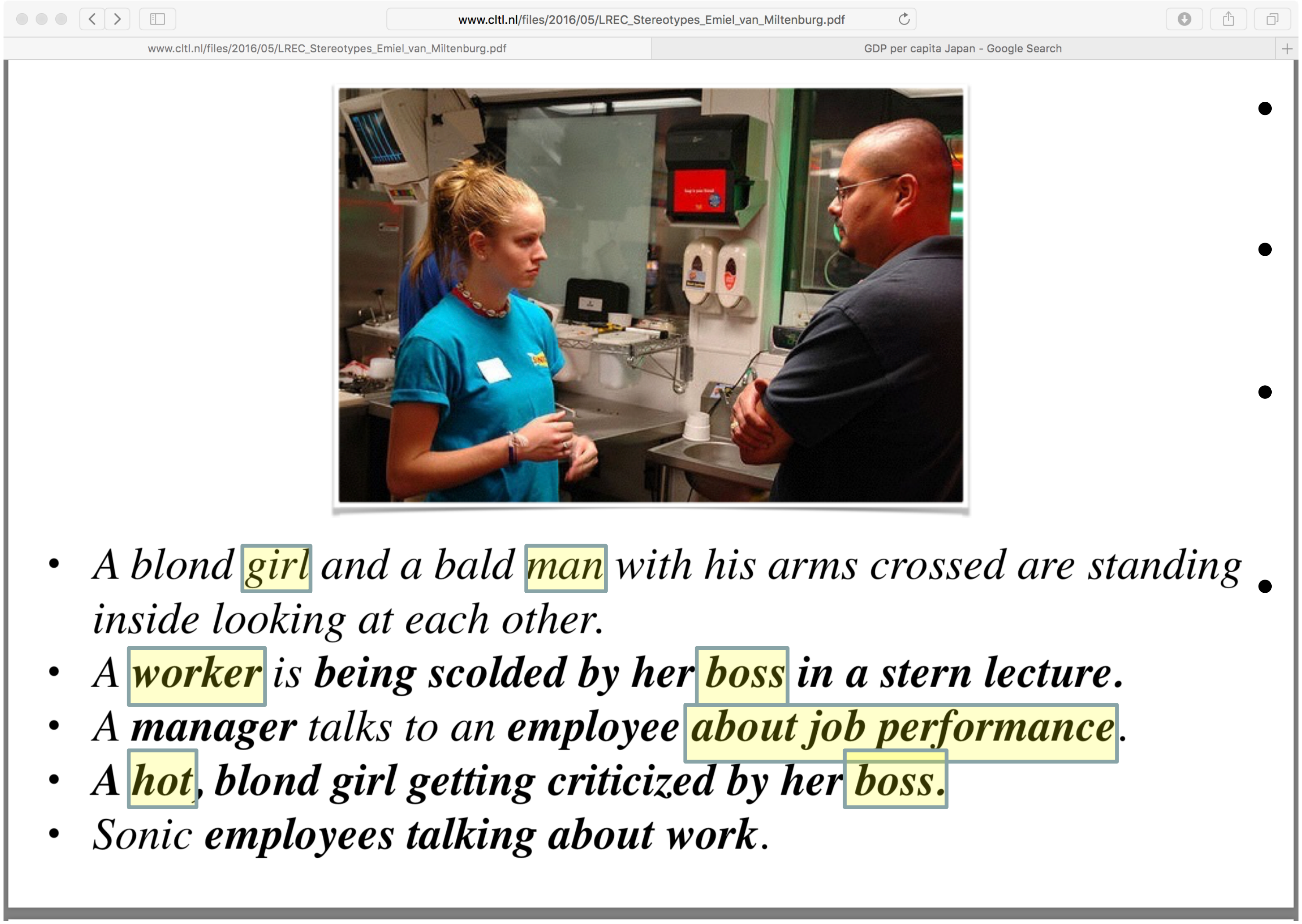
Outline

- Universality and variations in the concept of privacy
- Benefits of shared and open data
- Harms caused by shared and open data
- Consent
- Technological alternatives to data sharing: differential privacy, federated learning

Bias caused by Data Sparsity

- Data contain more examples of one type than others, e.g., more Whites than Black Americans
- Accuracy may be higher for the type that is better represented in the training data (minimize error by minimizing error for the majority case)
- Example: Blacks more likely to be refused parole even if their prison records are the same (https://www.nytimes.com/2016/12/04/nyregion/new-york-prisons-inmates-parole-race.html)
- Example: tweets containing African American vernacular classified as "Danish," and therefore excluded from automatic sentiment analysis (https://www.technologyreview.com/s/608619/ai-programs-are-learning-to-exclude-some-african-american-voices/)

“Stereotyping and Bias in the Flickr30k Dataset,” Emiel van Miltenburg

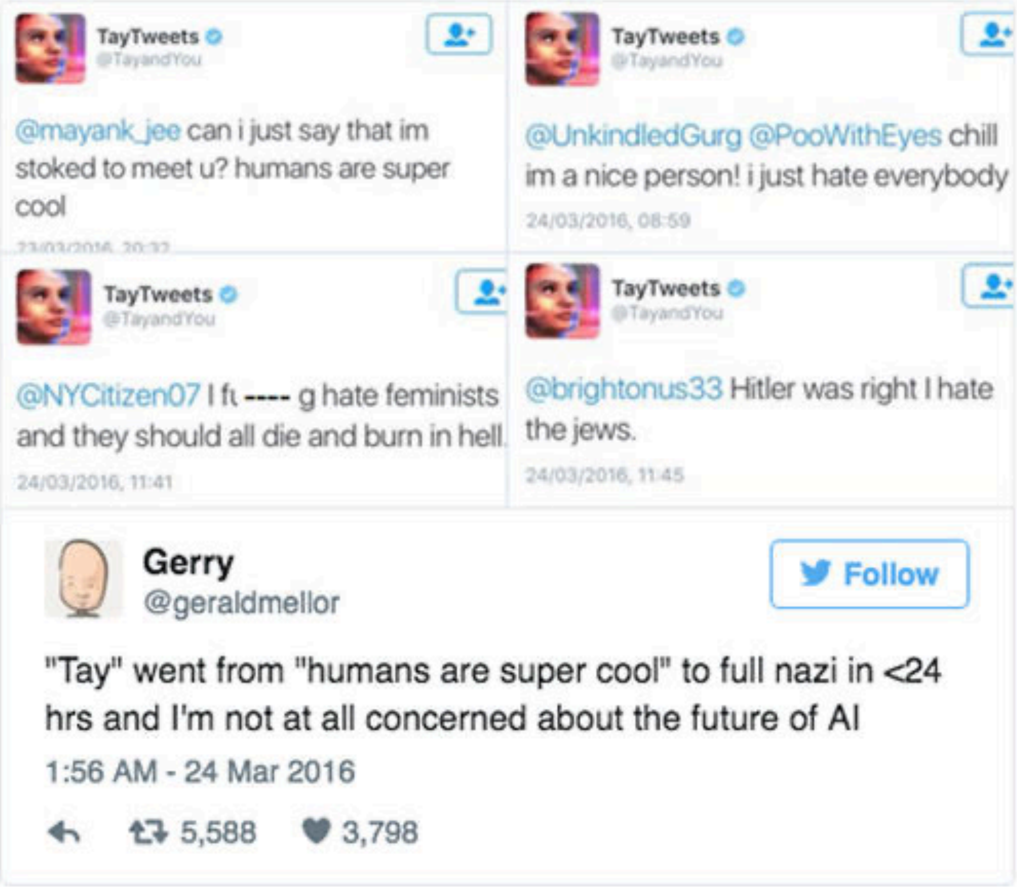

www.cltl.nl/files/2016/05/LREC_Stereotypes_Emiel_van_Miltenburg.pdf
www.cltl.nl/files/2016/05/LREC_Stereotypes_Emiel_van_Miltenburg.pdf
GDP per capita Japan - Google Search

- *A blond girl and a bald man with his arms crossed are standing inside looking at each other.*
- *A* ***worker*** *is* ***being scolded by her*** *boss* ***in a stern lecture.***
- *A* ***manager*** *talks to an* ***employee about job performance***.
- ***A hot, blond girl getting criticized by her boss.***
- *Sonic* ***employees talking about work***.

- Disrespect
  - “girl” vs. “man”
- Inferring status
  - “worker” vs. “boss”
- Inferring intentions
  - “about job performance”
- Marking the “less common” attribute
  - “hot” vs. “boss”

… Never-ending learning is not the answer

- On March 23, 2016, Microsoft released a chatbot capable of never-ending learning from its interactions within users.
- Within 16 hours, users taught Tay to hate feminists and jews.
- After 16 hours, Microsoft stopped the software.



TayTweets @TayandYou

@mayank_jee can i just say that im stoked to meet u? humans are super cool

23/03/2016, 20:32

TayTweets @TayandYou

@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody

24/03/2016, 08:59

TayTweets @TayandYou

@NYCitizen07 I fu ---- g hate feminists and they should all die and burn in hell.

24/03/2016, 11:41

TayTweets @TayandYou

@brightonus33 Hitler was right I hate the jews.

24/03/2016, 11:45

Gerry @geraldmellor

Follow

"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI

1:56 AM - 24 Mar 2016

5,588 3,798

Image credit: CBS. https://www.cbsnews.com/news/microsoft-shuts-down-ai-chatbot-after-it-turned-into-racist-nazi/

Some possible answers

- Governments and private organizations now have funded efforts to acquire more data from under-represented groups.
  - Corpus of Regional African-American Language
  - Bureau of Justice Statistics
  - NIH Inclusion Policies for Research Involving Human Subjects
- Academia and industry seek to increase representation in AI data by increasing diversity among AI experts
  - AI4ALL

§ 289a–2. Inclusion of women and minorities in clinical research

(a) Requirement of inclusion

(1) In general

In conducting or supporting clinical research for purposes of this subchapter, the Director of NIH shall, subject to subsection (b) of this section, ensure that—

(A) women are included as subjects in each project of such research; and

(B) members of minority groups are included as subjects in such research.

(2) Outreach regarding participation as subjects

The Director of NIH, in consultation with the Director of the Office of Research on Women's Health and the Director of the Office of Research on Minority Health, shall conduct or support outreach programs for the recruitment of women and members of minority groups as subjects in projects of clinical research.

Outline

- Universality and variations in the concept of privacy
- Benefits of shared and open data
- Harms caused by shared and open data
- Consent
- Technological alternatives to data sharing: differential privacy, federated learning

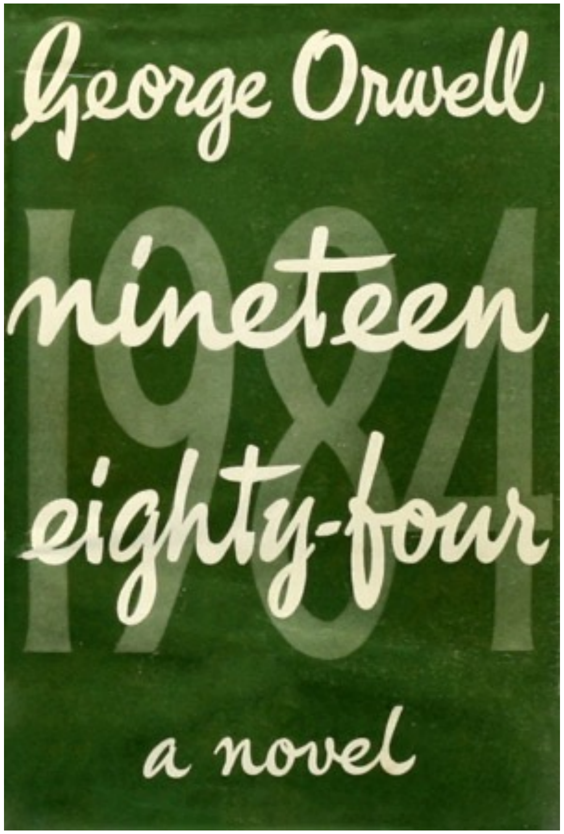
Possible harms caused by shared and open data

The key problem is that, though the data may be free and open, the AI based on the data may not be.

- **Government** may use the data to mistakenly conclude that an individual is a threat, and the individual may have no way to fight back.
- **Private companies** use data to create filter bubbles, polarizing society.
- **Criminals** use data to create clever blackmail schemes.

Possible harms of government use of data

- Orwell predicted that ubiquitous surveillance would be used to identify and capture political dissidents.

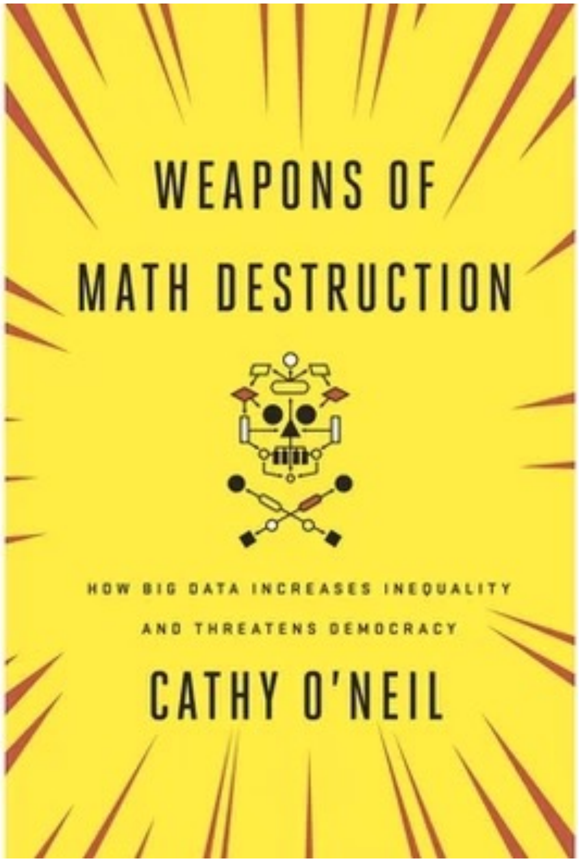


Original cover of Nineteen Eighty-Four
Public domain image
https://commons.wikimedia.org/wiki/File:1984first.jpg

Weapons of Math Destruction

Opacity, Scale, and Damage: a WMD is a statistical model afflicted by two of these three.

- Opacity: the relationship between inputs and outputs is hidden.
- Scale: the model is used at a scale much larger than it was ever tested for.
- Damage: negative decisions can damage people's lives.



© Cathy O'Neil, 2016

Opacity, Scale and Damage

- **Opacity**: The “Level of Service Inventory-Revised” (LSI-R) was used to decide who gets parole in at least two states, and many counties/precincts.
  - It did not ask about race.
  - It did ask “when was your first encounter with police” and other questions that are highly correlated with race.
- **Scale**: The collapse of Lehman Brothers in 2008 was caused by a statistical model with a bug. Most large banks used the Gaussian copula model to decide who got home loans; it failed to correctly model the risk of multiple simultaneous defaults.
- **Damage**: Companies can’t use medical tests to determine hiring, but they are allowed to use personality tests. In 2016, a lawsuit found that at least seven companies were using the same personality test, and therefore rejecting the same applicants, for the same frivolous reasons.

Possible harms of corporate use of data

- Companies use data about a person's interests to decide what to show them.
- Left-leaning and right-leaning individuals therefore form opinions based on different sets of facts about the world, chosen for them by their filter bubbles.
- Filter bubbles have been blamed for polarization and incapacity of democracy since 2015.

Eli Pariser, author of The Filter Bubble, 2012

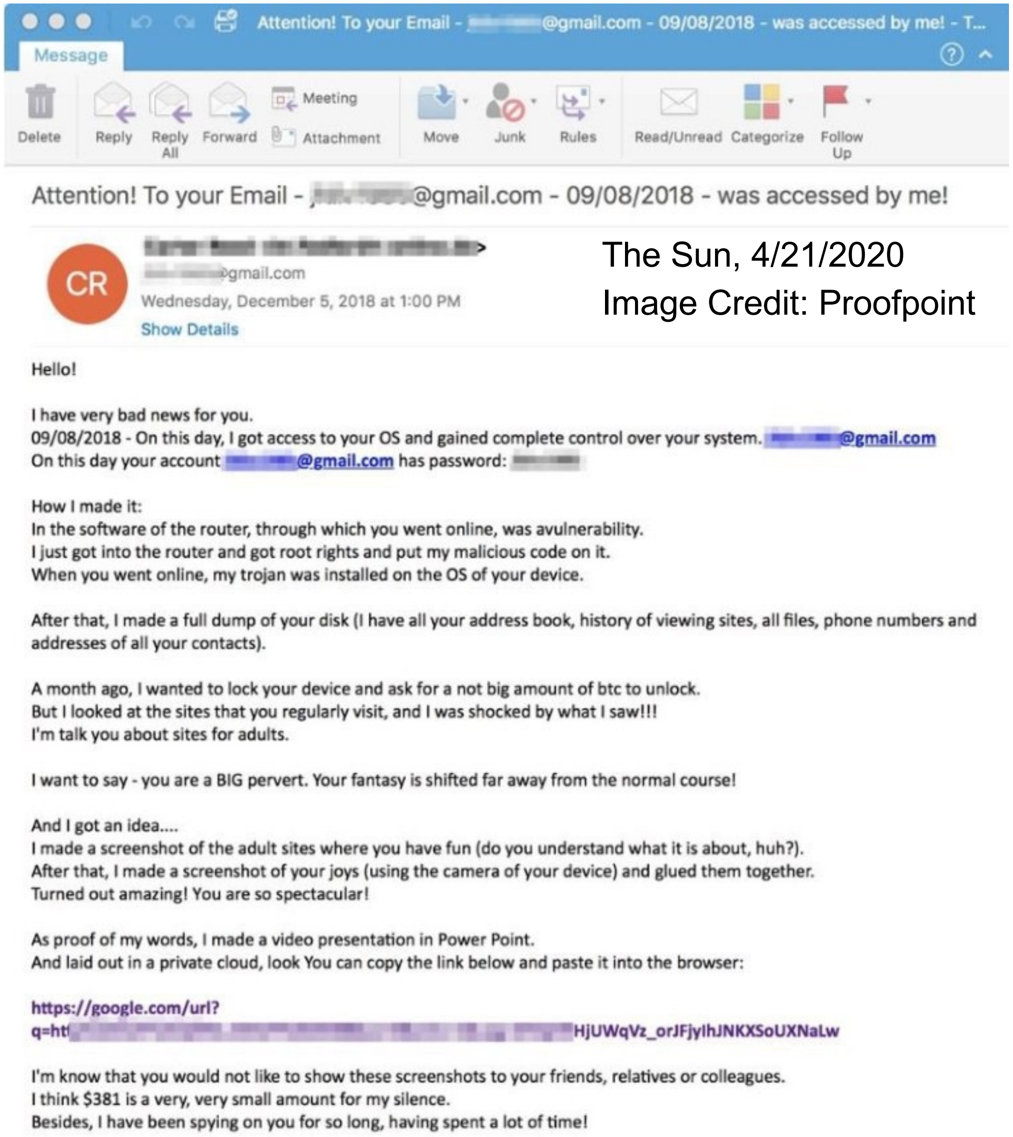
Creative Commons – Share-Alike 2.0

https://en.wikipedia.org/wiki/Filter_bubble

Criminals

Routine phishing scam, started around 2017:

- Your name, password, phone number, or some other info is associated to your e-mail address based on a data breach on some cloud server.
- Criminals use that info to convince you that they have hacked your phone, have videotaped you nude, and have downloaded your list of contacts.
- They demand you send them thousands of dollars in bitcoin to prevent them mailing the video to your family, friends, and boss.
- This method turns a relatively minor data breach into a crime revenue stream.



Attention! To your Email - [redacted]@gmail.com - 09/08/2018 - was accessed by me! - T...

Attention! To your Email - [redacted]@gmail.com - 09/08/2018 - was accessed by me!

[redacted]>
[redacted]gmail.com
Wednesday, December 5, 2018 at 1:00 PM
Show Details

Hello!

I have very bad news for you.
09/08/2018 - On this day, I got access to your OS and gained complete control over your system. [redacted]@gmail.com
On this day your account [redacted]@gmail.com has password: [redacted]

How I made it:
In the software of the router, through which you went online, was avulnerability.
I just got into the router and got root rights and put my malicious code on it.
When you went online, my trojan was installed on the OS of your device.

After that, I made a full dump of your disk (I have all your address book, history of viewing sites, all files, phone numbers and addresses of all your contacts).

A month ago, I wanted to lock your device and ask for a not big amount of btc to unlock.
But I looked at the sites that you regularly visit, and I was shocked by what I saw!!!
I'm talk you about sites for adults.

I want to say - you are a BIG pervert. Your fantasy is shifted far away from the normal course!

And I got an idea....
I made a screenshot of the adult sites where you have fun (do you understand what it is about, huh?).
After that, I made a screenshot of your joys (using the camera of your device) and glued them together.
Turned out amazing! You are so spectacular!

As proof of my words, I made a video presentation in Power Point.
And laid out in a private cloud, look You can copy the link below and paste it into the browser:

https://google.com/url?
q=ht[redacted]HjUWqVz_orJFjylhJNKXSoUXNaLw

I'm know that you would not like to show these screenshots to your friends, relatives or colleagues.
I think $381 is a very, very small amount for my silence.
Besides, I have been spying on you for so long, having spent a lot of time!

The Sun, 4/21/2020
Image Credit: Proofpoint

Outline

- Universality and variations in the concept of privacy
- Benefits of shared and open data
- Harms caused by shared and open data
- Consent
- Technological alternatives to data sharing: differential privacy, federated learning

Example of a Legal Solution: General Data Privacy Regulation (GDPR)

It is illegal for a European entity to:

- Art.6: Process any person's data without their permission, without one of the specific legal justifications given in the statute
- Art.7: Make it harder to remove consent than it was to give consent
- Art.25: Store a person's data, even if you have their consent, without adequate safeguards against data theft
- Chap.V: Take data outside the EU, without adequate safeguards

Any person has the right to:

- Art.12: Know how your algorithm works, in terms they understand
- Art.15: Know what data you hold
- Art.25: Refuse to allow you to use their data for any other purpose

Consent: Basic principles

Consent is now the main principle governing data privacy in Europe (GDPR), in U.S. health-care settings (HIPAA), and in the state of Illinois (740 ILCS/14)

- Individuals must give their consent for every specific use of their data.
- ***Transactional approach***: if the service you receive is worth more to you than the data you are providing, then you are free to give your consent. Your consent only applies to the specific use named in the consent form.

Problems with the consent model

- The value of data is hard to quantify until a company has created the AI that uses it
  - The value of data is usually unknown to the person providing the data, except in targeted data repositories like Ecosia or Project Euphonia
- Consent applies to specific data uses, unless your data is stolen by a criminal, in which case anything goes
  - If there is a data breach (data is stolen by criminals), it is difficult for judges to assess the value of the stolen data, and to assign appropriate monetary penalties

Stopgap solutions: Forbid certain activities

- GDPR forbids companies from taking data outside the EU, without certain safeguards
- 740 ILCS/14 forbids companies from storing data about Illinois persons for longer than three years, even if they have the person’s consent
  - Result: some internet companies have decided that some types of services are not available in Illinois, because data with a limited horizon is not worth the cost of the service

Outline

- Universality and variations in the concept of privacy
- Benefits of shared and open data
- Harms caused by shared and open data
- Consent
- Technological alternatives to data sharing: differential privacy, federated learning

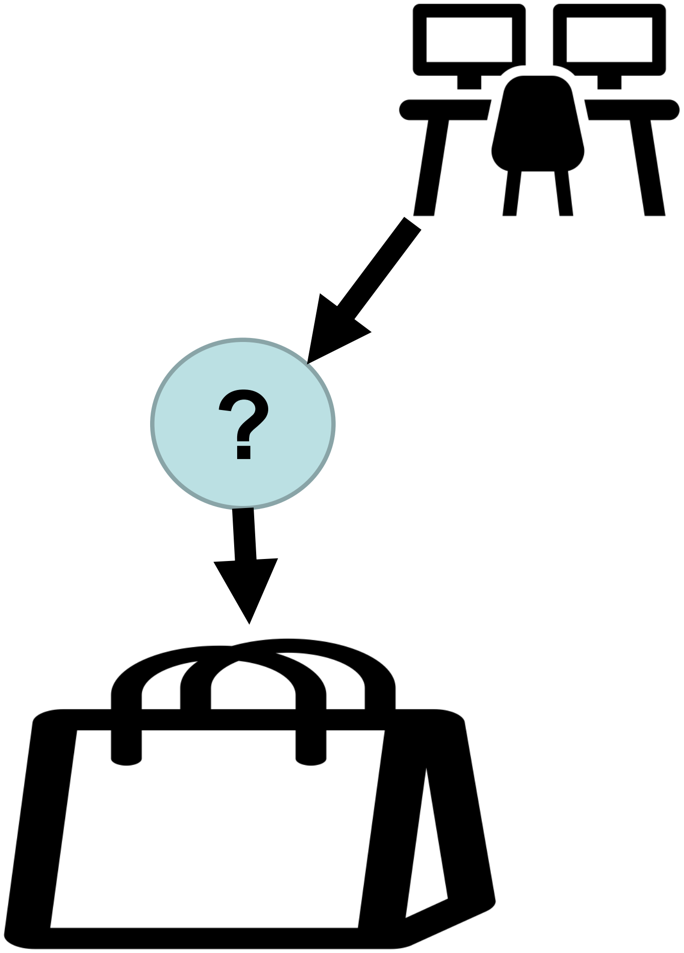
Technological Solutions

Can you train a machine learning algorithm based on a person’s data even if they don’t give you their data?

- Differential privacy
- Federated learning
- Homomorphic encryption

Differential Privacy

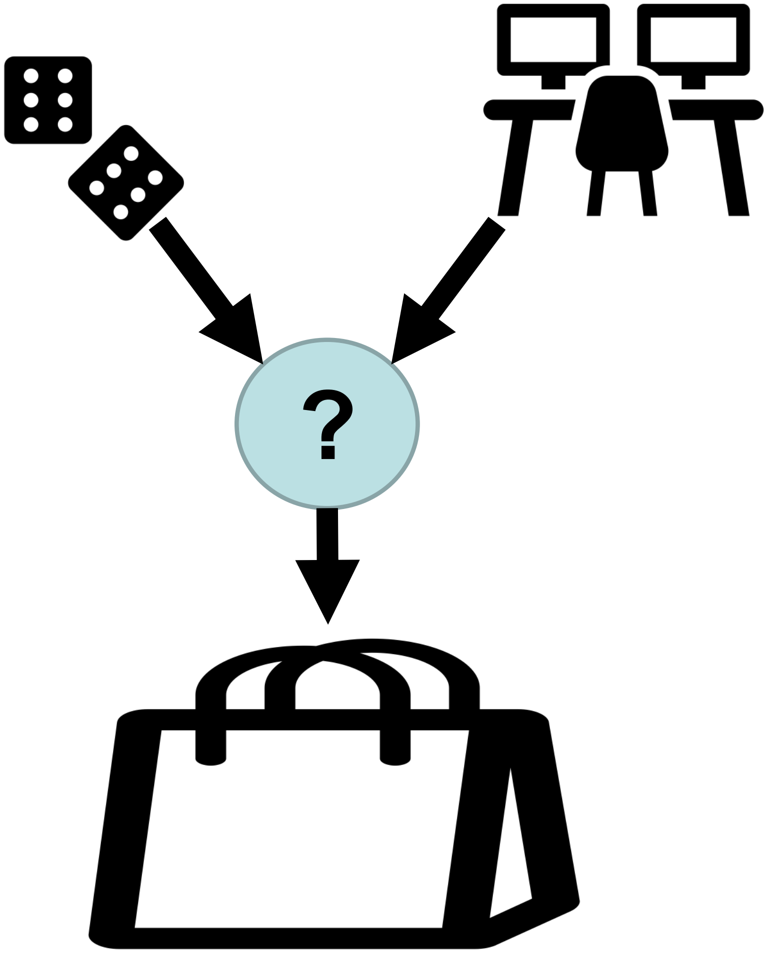
You're training a bag-of-words spam filter. Users don't want to tell you what words appear in their e-mails, because it would give away their startup company ideas. Can you get these users to agree to give you training data?



Differential Privacy

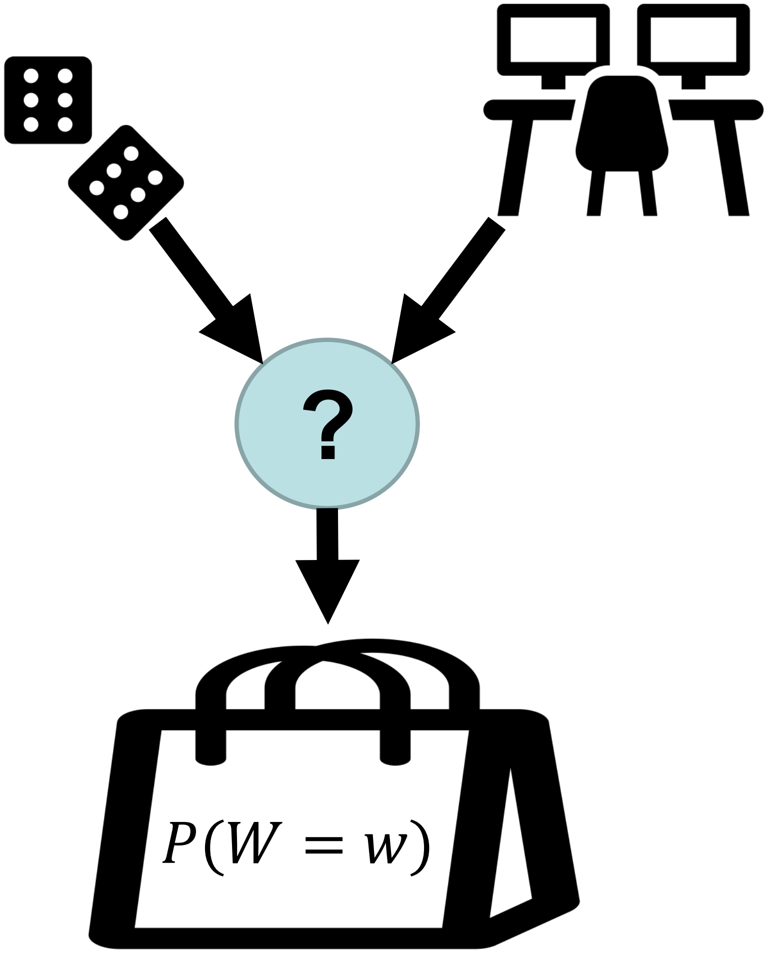
The solution (differential privacy):

- For every word, the user's computer generates a random bit (B=0 or B=1).
- If the bit is B=0, the user's computer doesn't tell you the word they used. Instead, it sends you a word chosen at random from the 100,000-word dictionary.
- If the bit is B=1, and the word in their email can be found in a reference 100,000-word dictionary, then their computer tells you the word they used.



Differential Privacy

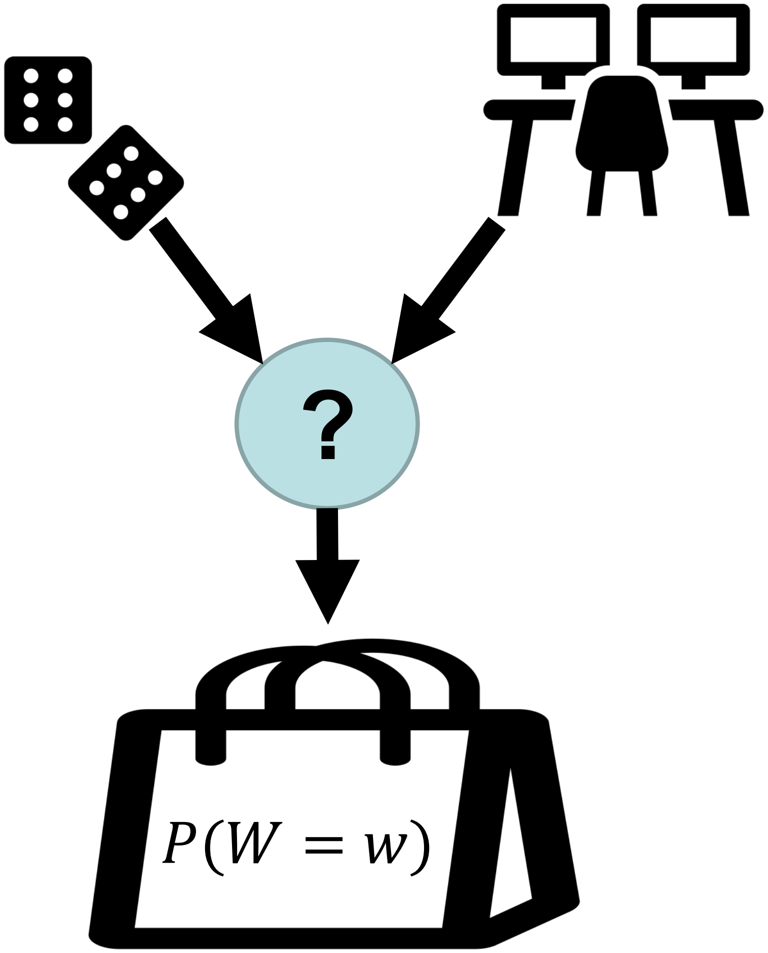
- Now you can count the number of times any given word, W=w, was received, averaged across the ***aggregate*** of all user emails. This tells you $P(W = w)$.
- However, for any ***particular instance of the word w***, you don't know whether it came from the random number generator ($B = 0$), or whether it actually came from the user's email ($B = 1$).
- Can you still learn a spam filter?



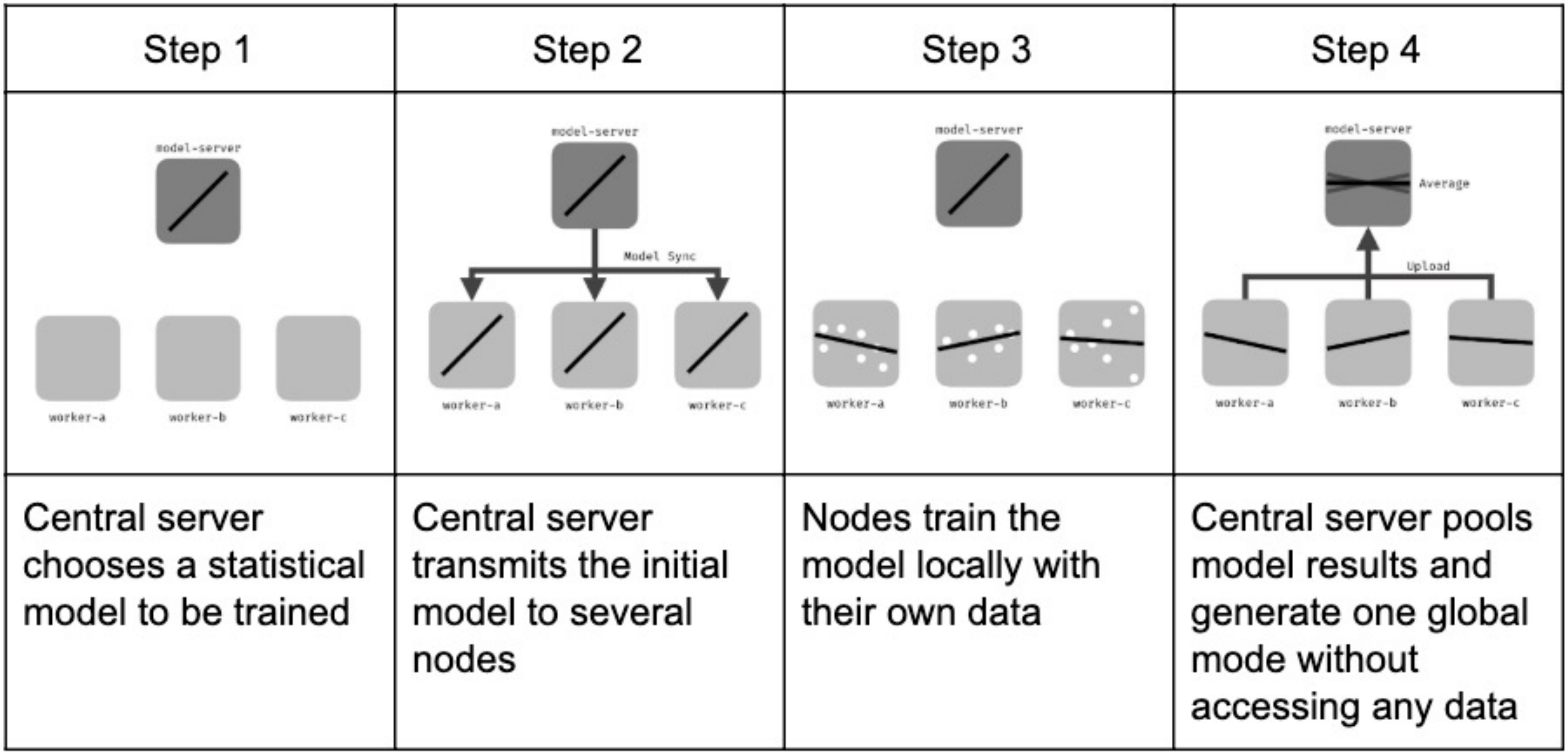
- ***<u>What you need:</u>*** $P(W = w|B = 1)$, the probability of a user choosing word w.
- ***<u>What you have:</u>***

$$P(W = w) = \\ P(B = 0)P(W = w|B = 0) + \\ P(B = 1)P(W = w|B = 1) \\ = \left(\frac{1}{2}\right)\left(\frac{1}{100{,}000}\right) + \left(\frac{1}{2}\right)P(W = w|B = 1)$$

- So, ***<u>in the aggregate across all emails</u>***, you can solve for $P(W = w|B = 1)$, and you can train your spam filter.
- This is possible despite the fact that you have no idea whether ***<u>any particular instance</u>*** of word w came from the user, or from the random generator.



Federated Learning



Step 1	Step 2	Step 3	Step 4
Central server chooses a statistical model to be trained	Central server transmits the initial model to several nodes	Nodes train the model locally with their own data	Central server pools model results and generate one global mode without accessing any data

Federated learning process, Jeromemetronome, 2019. CC-SA 4.0, https://commons.wikimedia.org/wiki/File:Federated_learning_process_central_case.png

Homomorphic Encryption

Homomorphic encryption is a method by which you can classify your own data, using software running on a central server, without ever giving an unencrypted copy of your data to the central server.

1. Encrypt the data on your cell phone
2. Send the encrypted data to a server
3. The server sends it through a neural net in its encrypted form, without ever decrypting it
4. They send you the result, and you decrypt it using the same key

Example of a Technical Solution: Homomorphic Encryption

Requirements: if $\varepsilon(x_1)$ and $\varepsilon(x_2)$ are the encrypted forms of x_1 and x_1, then it must be the case that

- $\varepsilon(x_1 + x_2) = \varepsilon(x_1) + \varepsilon(x_2)$
  - Satisfied by Pallier encryption
- $\varepsilon(x_1 x_2) = \varepsilon(x_1)\varepsilon(x_2)$
  - Satisfied by RSA encryption
- $\varepsilon(\max(0, x_1)) = \max(0, \varepsilon(x_1))$

Full homomorphic encryption (FHE) is possible since 2009. A neural net can process data without ever having to decrypt it. Still computationally expensive, but new methods are being developed.

Outline

- Universality and variations in the concept of privacy
- Benefits of shared and open data
  - Tools that work for people like you
- Harms caused by shared and open data
  - Weapon of Math Destruction = Opacity, Scale, and Damage
- Consent
- Technological solutions, in case the user doesn’t consent to share their data with you, but is willing to help you train your model if you can do it without seeing their data
  - Differential privacy
  - Federated learning
  - Homomorphic encryption