CS440/ECE448 Lecture 25: Privacy

Mark Hasegawa-Johnson, 3/2024

"One nation under CCTV," Banksy, 2008. CC-SA 2.0, https://commons.wikimedia.org/wiki/File:Bansky_one_nation_under_cctv.jpg

Outline

- Legal framework: Harm
- Algorithms: 1 that fails, 1 that works, 2 that might work
  - Binning/K-anonymity
  - Differential privacy
  - Federated learning
  - Homomorphic encryption
- Legal framework: Rights
- Rights-based algorithmic framework

A brief history of privacy

- In a rural society, you can be alone whenever you want (walk out into the field)…
- …but being *completely* alone is dangerous (animals, bandits).
- Many societies developed activities or places that could be accessed by a limited group.

https://commons.wikimedia.org/wiki/File:Uluru_(Helicopter_view)-crop.jpg

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.

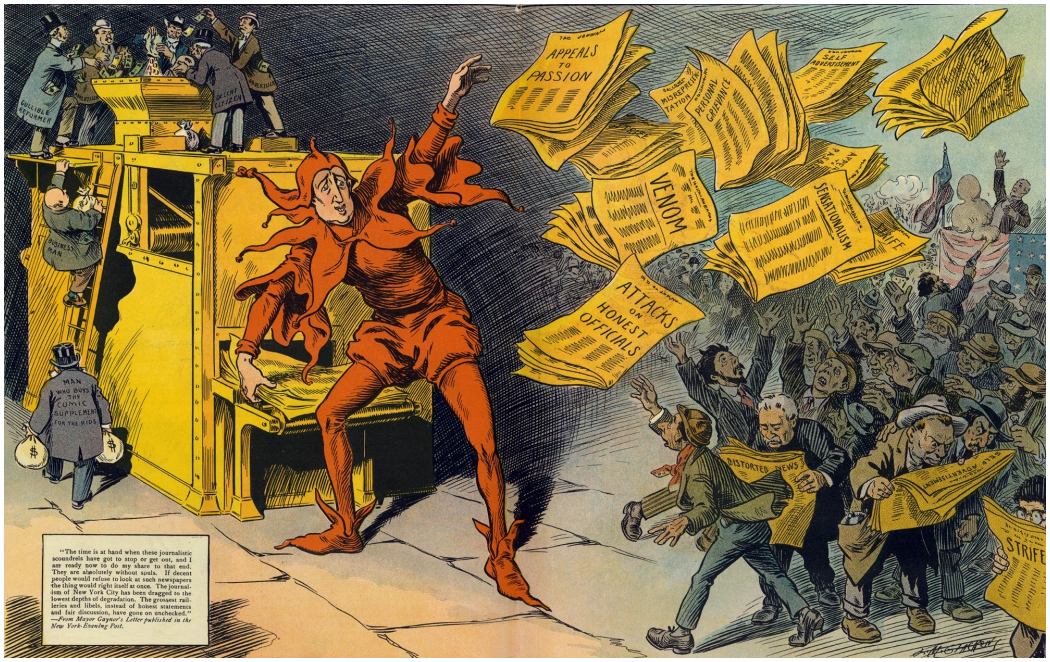
A brief history of privacy

- In pre-industrial urban civilizations, being alone means that you have to do all of your own work.
- People of high social status were *never* alone.
- In order to be around a person of high social status, therefore, you had to agree to follow strict social protocol, granting them control of their environment.

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

A brief history of privacy

- In 19th-century UK and US, newspapers stopped deferring to high social status; they became popular by printing gossip.
- People of high social status felt attacked.
- High-status people invented the idea of "the right of privacy" as a kind of self-defense.
- In an 1890 article, Warren and Brandeis defined defined "The Right to Privacy" as "the right to be left alone."

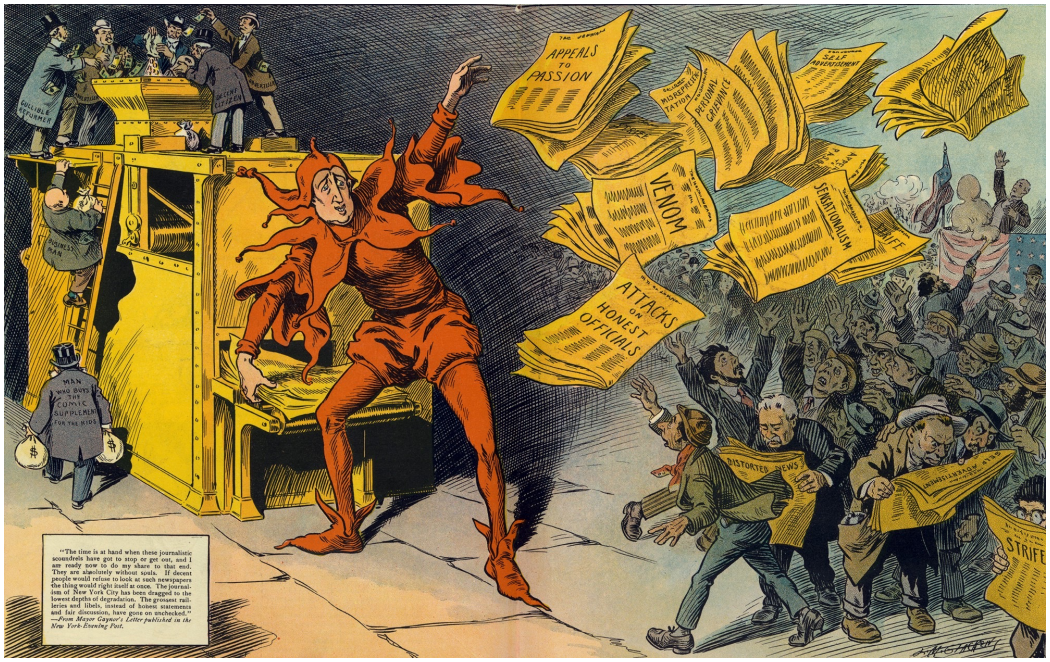


The Yellow Press, by L.M. Glackens, 1910.

Public domain image, https://commons.wikimedia.org/wiki/File:The_Yellow_Press_by_L.M._Glackens.jpg

Legal framework: Harms

- Most privacy laws in the U.S. are still based on protection from harm.
- If you can demonstrate that somebody harmed you by publishing something private about you, then you can sue them.



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

Responsibilities of data scientists under a harms-based legal framework

- Under a harms-based legal framework, those of us who work with large amounts of data are responsible for ensuring that our processing of the data does not cause harm.
- The most frequent type of harm is data theft, therefore most algorithms are focused on preventing data theft.

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

Note: Users can be harmed if their data is NOT available to help train AI.

- Koenecke et al.(doi:10.1073/pnas.1915768117, 2020) tested automatic speech recognition software published by Amazon, Apple, Google, IBM and Microsoft
- Data: autobiographical monologs by black (73) and white (42) people
- Result: word error rate was 35% for black speakers, 19% for white speakers
- Why:
  - Training data includes more white people than black people.
  - The variability in the speaking styles of different white people is well-represented in training data, but the variability in speaking styles of different black people is not well-represented.

Outline

- Legal framework: Harm
- Algorithms: 1 that fails, 1 that works, 2 that might work
  - Binning/K-anonymity
  - Differential privacy
  - Federated learning
  - Homomorphic encryption
- Legal framework: Rights
- Rights-based algorithmic framework

Binning/ k-anonymity

- In order to make people anonymous, can we just remove their name and religion?
- …no, because there is only one 28-year-old in the table, so even without her name, we know who she is.
- Proposed solution: report age in 10-year increments. Now there are four women in the 20-30 bin, so their privacy is kind of protected.

Name	Age	Gender	Height	Weight	State of domicile	Religion	Disease
Ramsha	30	Female	165cm	72kg	Tamil Nadu	Hindu	Cancer
Yadu	24	Female	162cm	70kg	Kerala	Hindu	Viral infection
Salima	28	Female	170cm	68kg	Tamil Nadu	Muslim	Tuberculosis
Sunny	27	Male	170cm	75kg	Karnataka	Parsi	No illness
Joan	24	Female	165cm	71kg	Kerala	Christian	Heart-related
Bahuksana	23	Male	160cm	69kg	Karnataka	Buddhist	Tuberculosis
Rambha	19	Male	167cm	85kg	Kerala	Hindu	Cancer
Kishor	29	Male	180cm	81kg	Karnataka	Hindu	Heart-related
Johnson	17	Male	175cm	79kg	Kerala	Christian	Heart-related

Name	Age	Gender	Height	Weight	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	165cm	72kg	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	162cm	70kg	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	170cm	68kg	Tamil Nadu	*	Tuberculosis
*	20 < Age ≤ 30	Male	170cm	75kg	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	165cm	71kg	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male	160cm	69kg	Karnataka	*	Tuberculosis
*	Age ≤ 20	Male	167cm	85kg	Kerala	*	Cancer
*	20 < Age ≤ 30	Male	180cm	81kg	Karnataka	*	Heart-related
*	Age ≤ 20	Male	175cm	79kg	Kerala	*	Heart-related
*	Age ≤ 20	Male	169cm	82kg	Kerala	*	Viral infection

Binning/ k-anonymity

- In order to guarantee that binning works, we need to guarantee that the provided data never specifies one individual exactly.
- A **k-anonymous** binning algorithm guarantees that the provided data will always refer to at least k different individuals.
- Unfortunately, no guaranteed K-anonymizing algorithms exist. Many datasets that seem to be K-anonymized have been successfully de-anonymized.

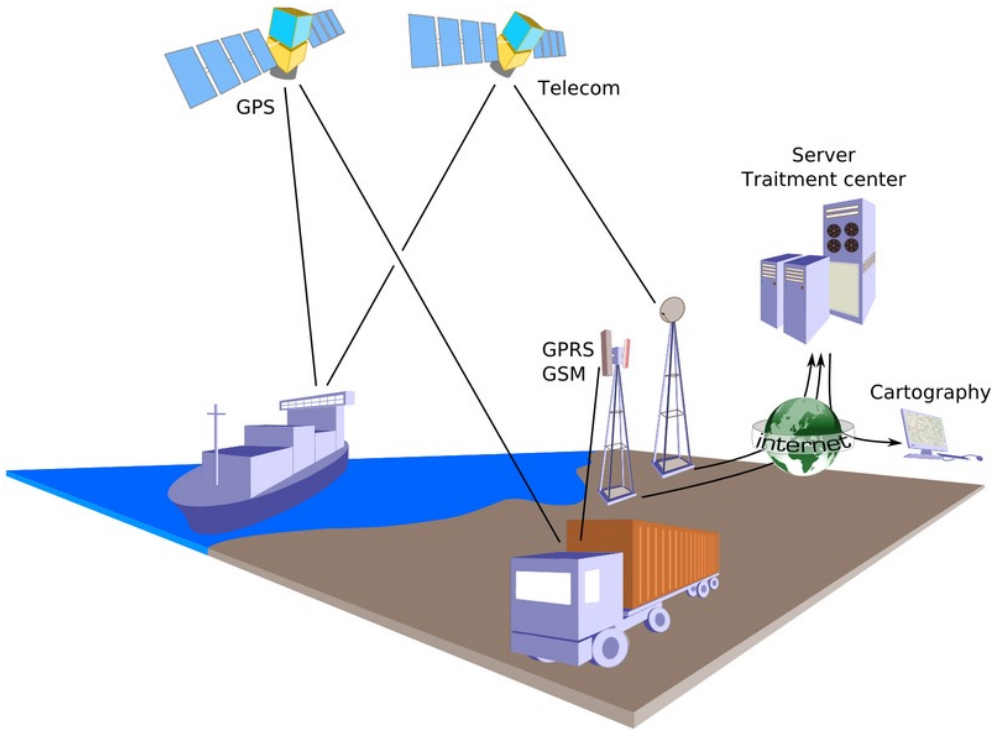
Name	Age	Gender	Height	Weight	State of domicile	Religion	Disease
Ramsha	30	Female	165cm	72kg	Tamil Nadu	Hindu	Cancer
Yadu	24	Female	162cm	70kg	Kerala	Hindu	Viral infection
Salima	28	Female	170cm	68kg	Tamil Nadu	Muslim	Tuberculosis
Sunny	27	Male	170cm	75kg	Karnataka	Parsi	No illness
Joan	24	Female	165cm	71kg	Kerala	Christian	Heart-related
Bahuksana	23	Male	160cm	69kg	Karnataka	Buddhist	Tuberculosis
Rambha	19	Male	167cm	85kg	Kerala	Hindu	Cancer
Kishor	29	Male	180cm	81kg	Karnataka	Hindu	Heart-related
Johnson	17	Male	175cm	79kg	Kerala	Christian	Heart-related

Name	Age	Gender	Height	Weight	State of domicile	Religion	Disease
*	20 < Age ≤ 30	Female	165cm	72kg	Tamil Nadu	*	Cancer
*	20 < Age ≤ 30	Female	162cm	70kg	Kerala	*	Viral infection
*	20 < Age ≤ 30	Female	170cm	68kg	Tamil Nadu	*	Tuberculosis
*	20 < Age ≤ 30	Male	170cm	75kg	Karnataka	*	No illness
*	20 < Age ≤ 30	Female	165cm	71kg	Kerala	*	Heart-related
*	20 < Age ≤ 30	Male	160cm	69kg	Karnataka	*	Tuberculosis
*	Age ≤ 20	Male	167cm	85kg	Kerala	*	Cancer
*	20 < Age ≤ 30	Male	180cm	81kg	Karnataka	*	Heart-related
*	Age ≤ 20	Male	175cm	79kg	Kerala	*	Heart-related
*	Age ≤ 20	Male	169cm	82kg	Kerala	*	Viral infection

Example: Geolocation data

Montjoye et al. (doi:10.1038/srep01376, 2013) showed that,

- In a database of 1.5 million people,
- given the ID # of the cell-phone base station closest to a user at 4 different times,
- it is possible to uniquely identify 95% of all users.



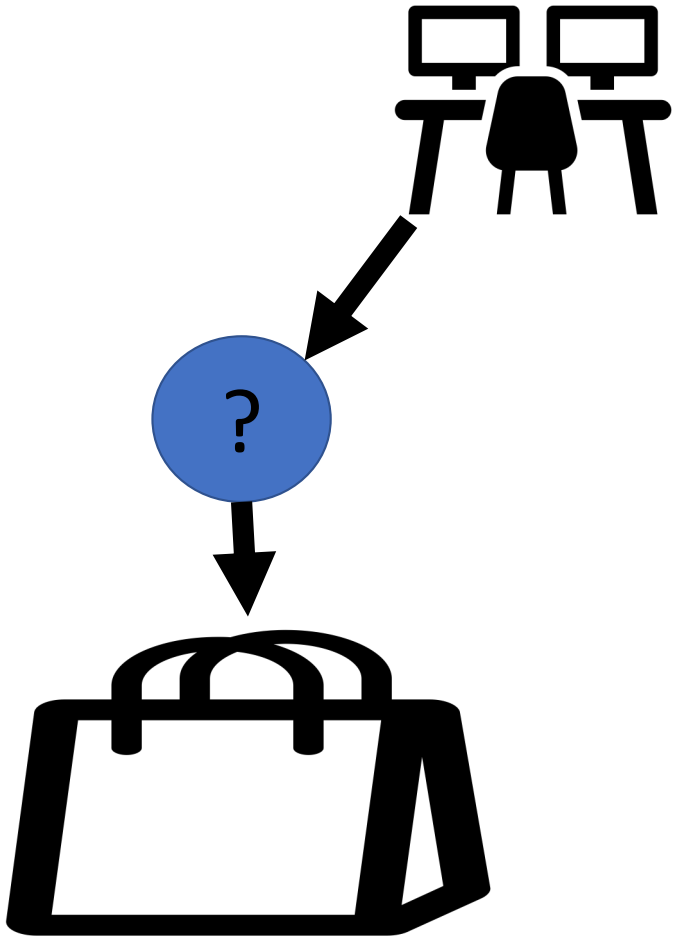
GFDL, Éric Chassaing, https://commons.wikimedia.org/wiki/File:Geolocation.png

Outline

- Legal framework: Harm
- Algorithms: 1 that fails, 1 that works, 2 that might work
  - Binning/K-anonymity
  - Differential privacy
  - Federated learning
  - Homomorphic encryption
- Legal framework: Rights
- Rights-based algorithmic framework

Differential Privacy

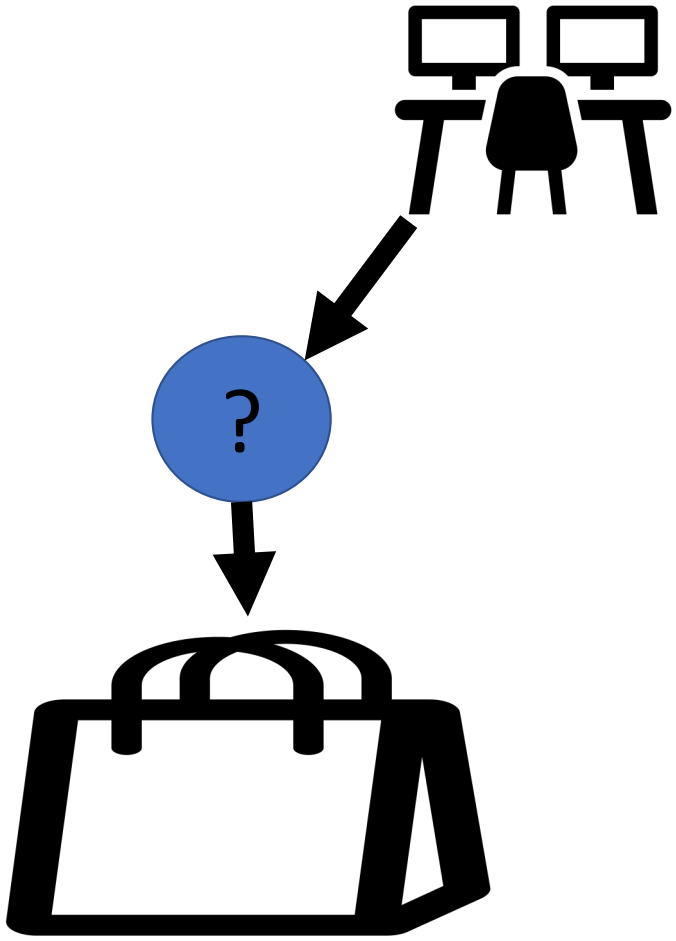
You're training a bag-of-words spam filter. Users don't want to tell you what words appear in their e-mails, because it would give away their startup company ideas. Can you get these users to agree to give you training data?



Differential Privacy

The goal:

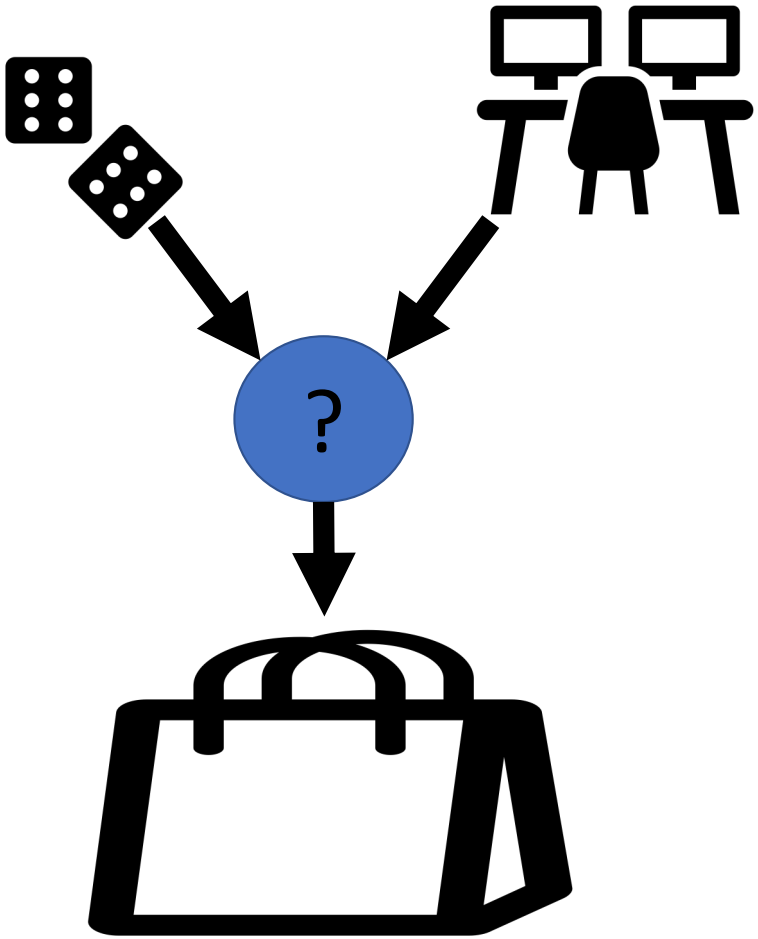
- Let Y = the true word. You want to accurately estimate $p = \Pr(Y = \text{coffee})$, the frequency of the word “coffee” in real-world e-mails that people send to each other.
- You don’t want to know whether any given user’s e-mail contains the word ”coffee.”



Differential Privacy

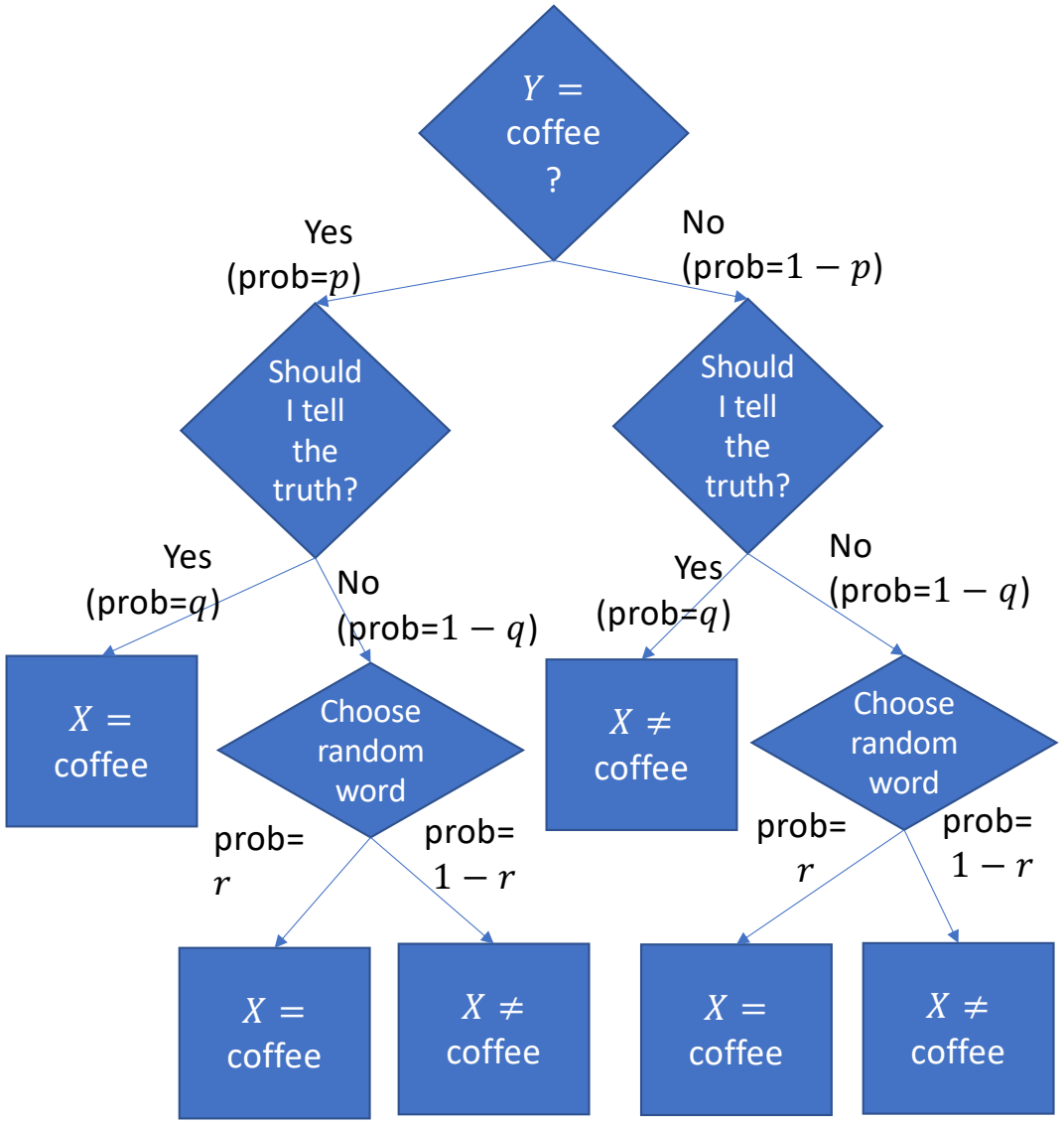
The solution (differential privacy):

- Let X = the word the user reports to you
- With probability q, the user tells you the word they really used ($X = Y$).
- With probability $1 - q$, the user lies. They choose some word at random from a $(1/r)$-word dictionary, and tell you that word instead.



Differential Privacy

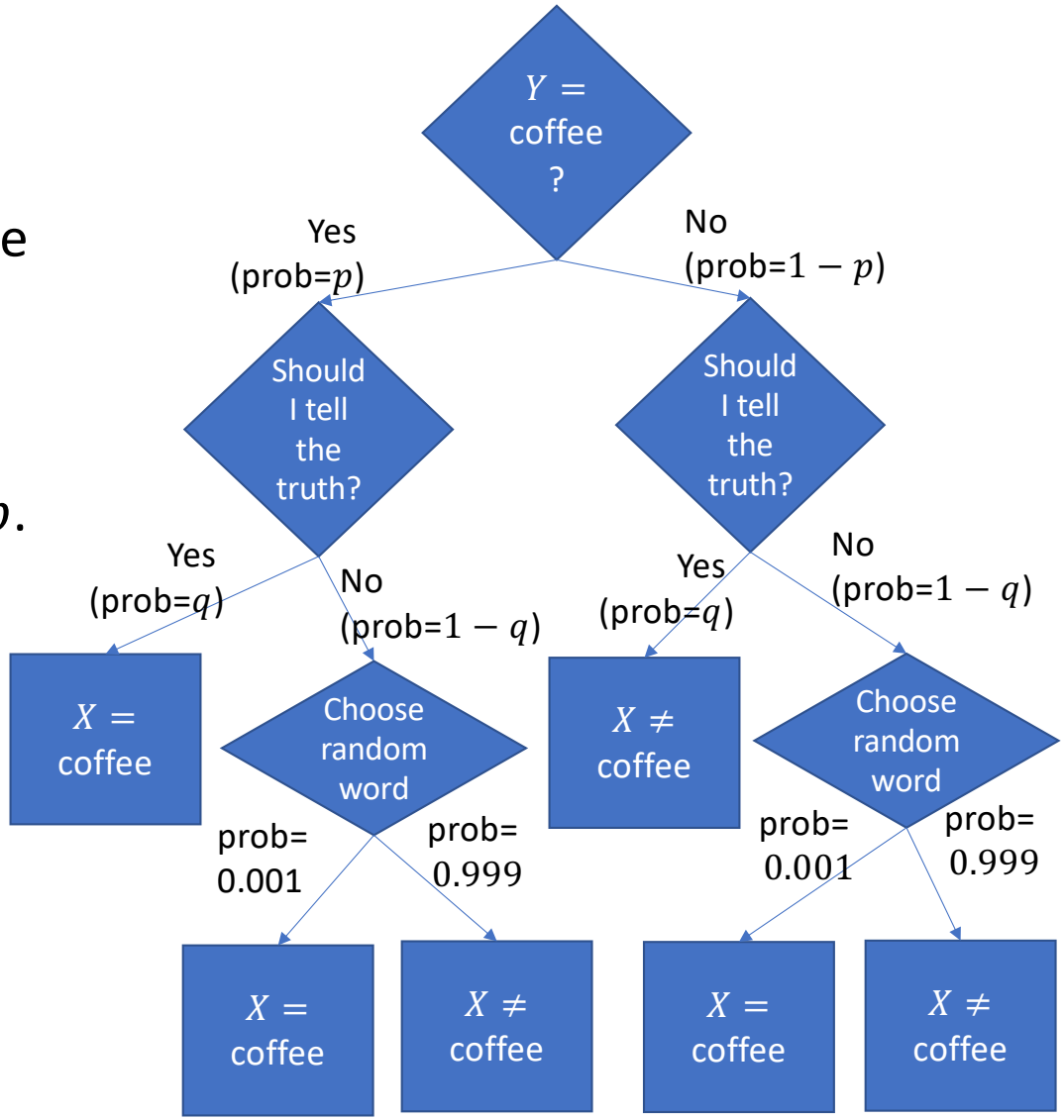
- $p = \Pr(Y = \text{coffee})$
  - Probability that the next word in the user's e-mail is "coffee"
- $q = \Pr(X = Y)$
  - Probability that the user tells you the truth
- $r = \Pr(X = \text{coffee} | X \neq Y)$
  - Probability of selecting "coffee" at random from the dictionary



Differential Privacy

- Frequency of the word “coffee” in the data that users send you:
  $\Pr(X = \text{coffee}) = pq + r(1-q)$
- Since you know q and r, you can easily calculate the correct value of p.
- …but you have no idea whether any particular person used the word “coffee:”

$$\Pr(Y = \text{coffee} | X = \text{coffee}) = \frac{pq}{pq + r(1-q)}$$



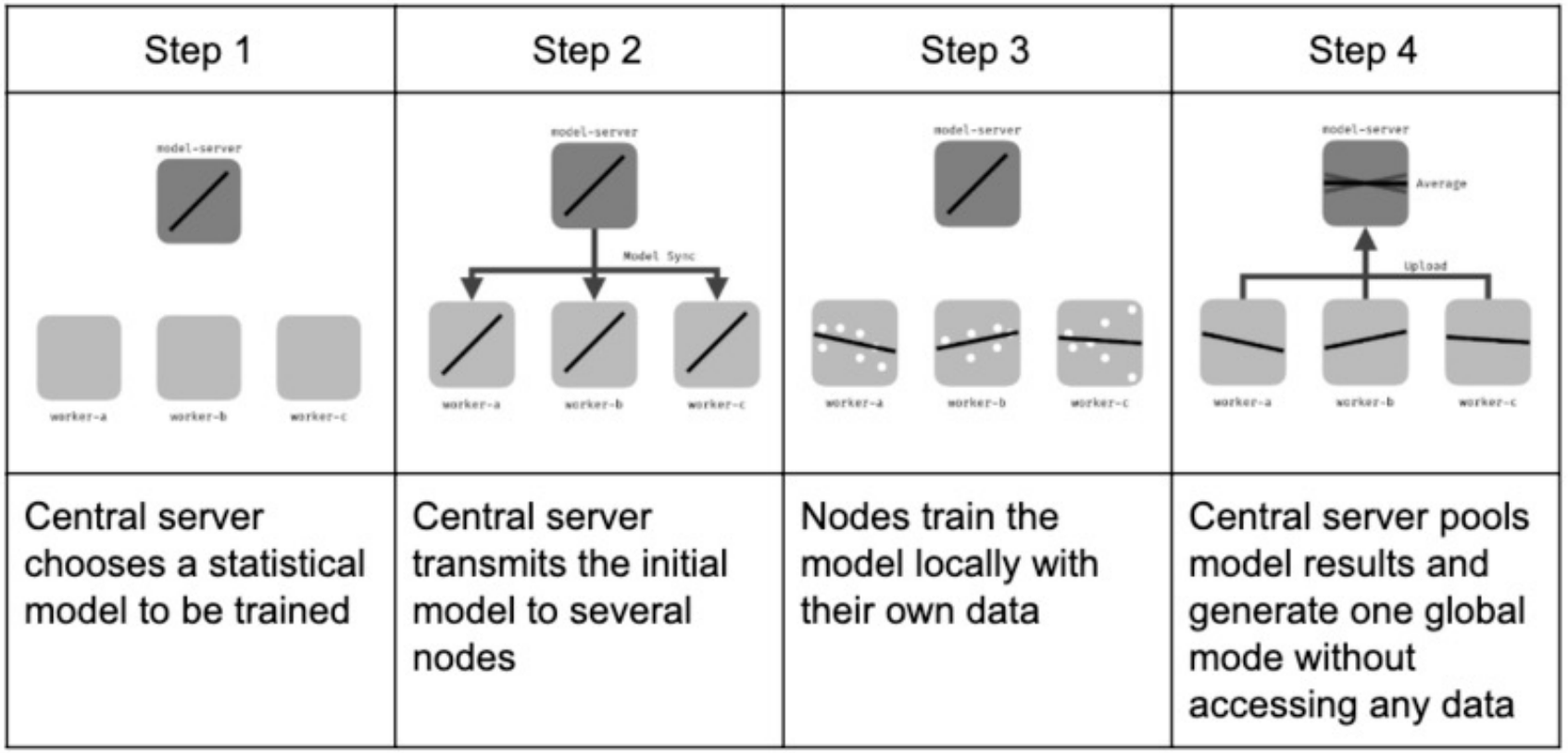
Quiz

Try the quiz:
https://us.prairielearn.com/pl/course_instance/147925/assessment/2409479

Outline

- Legal framework: Harm
- Algorithms: 1 that fails, 1 that works, 2 that might work
  - Binning/K-anonymity
  - Differential privacy
  - Federated learning
  - Homomorphic encryption
- Legal framework: Rights
- Rights-based algorithmic framework

Federated learning



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

Does Federated Learning work?

- Some studies have shown that you can reconstruct training data from a fully-trained neural net, so even if users only send you the trained model parameters, you can reconstruct their data
- Current research: find training algorithms based on differential privacy, so that the network learns without telling you its data

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 processes the encrypted data, in order to generate an encrypted answer, which is returned to you
4. You decrypt the answer on your cell phone

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

- Legal framework: Harm
- Algorithms: 1 that fails, 1 that works, 2 that might work
  - Binning/K-anonymity
  - Differential privacy
  - Federated learning
  - Homomorphic encryption
- Legal framework: Rights
- Rights-based algorithmic framework

Legal framework: Rights

- A rights-based framework defines ”control over your own data” as a human right
- You can sue people who violate your rights, even if they did not harm you

General Data Privacy Regulation (GDPR)

Europeans have the right to:

- Learn where their data is stored, and access to it
- Have their data stored in a manner that prevents unauthorized release
- Correct their data if there are mistakes
- Object to processing of their data, using a binary option that is clearly described and that does not try to hide the “no” option

Data may not be transferred to other countries or international organizations unless the EU has determined that the recipient has adequate data privacy safeguards.

GDPR violations may be fined up to 2% of your global gross revenue!

Problems with GDPR

- If data is stolen: Under what circumstance is the data processing company responsible? What types of anti-theft safeguards are considered legally sufficient?
- What is a “clearly described” binary option? How clearly visible does the “no” option need to be?

Disagreement about these questions have led to some of the largest lawsuits in the history of the world, since the penalty can be up to 2% of a company’s gross worldwide revenue.

Illinois Biometric Information Privacy Act: An example of a badly-written law

740 ILCS 14/15: An individual or company can hold biometric data (voice, face) of any person living in Illinois only if:

a) They have a written policy

b) They have obtained your consent

c) They do not profit from it

d) They don't give it away without your consent

e) They protect it from data theft

If any of the above is violated, you can sue them, even if the violation didn't hurt you.

(a),(b),(d),(e) = GDPR-like right to privacy

(c) = specifically prevents the inclusion of your data in commercial AI training algorithms, which can cause harm to the user.

Outline

- Legal framework: Harm
- Algorithms: 1 that fails, 1 that works, 2 that might work
  - Binning/K-anonymity
  - Differential privacy
  - Federated learning
  - Homomorphic encryption
- Legal framework: Rights
- Rights-based algorithmic framework

Rights-based algorithmic framework

In “A Bold New Plan for Preserving Online Privacy and Security” (IEEE Spectrum, 12/2023), Raghavan & Schneier propose:

- All of your data online is encrypted, and only you have the key.
  - Your cloud service provider never knows your unencrypted data.
- If Acme Corporation needs your data for something:
  - You send the encrypted data to Acme.
  - You grant Acme permission to use your data, for a specified use, by decrypting it in the app where Acme is using it.

Example

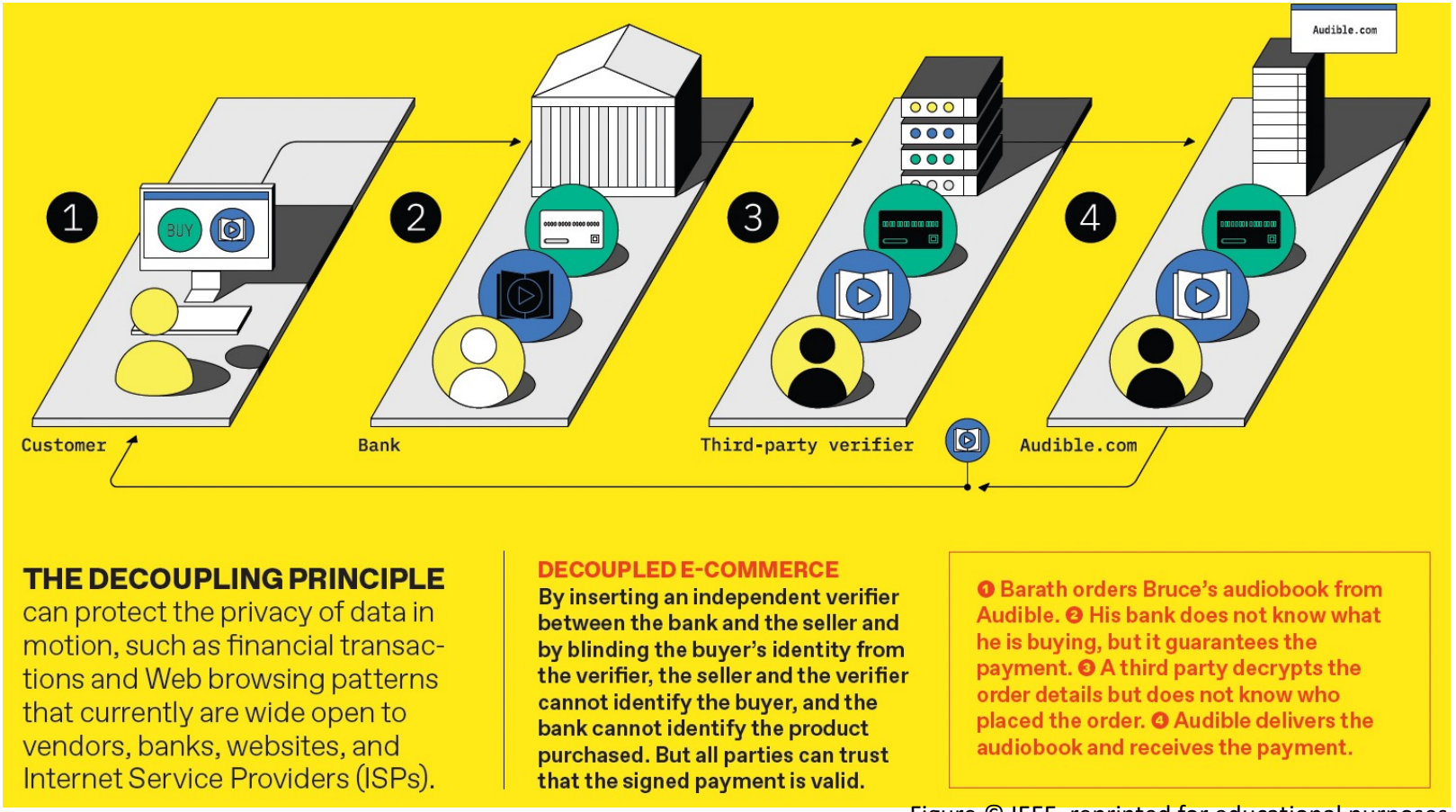


Figure © IEEE, reprinted for educational purposes

- Audible.com knows you and the verifier, but doesn't know your bank.
- Your bank knows you and the verifier, but doesn't know that you purchased from Audible.com.
- Verifier knows that somebody at your bank purchased something from Audible.com, but doesn't know it was you, or what it was that you purchased.

Machine learning under a data-rights framework

- The encrypt-until-used framework works well with a federated learning framework incorporating differential privacy. Participant learns about a new AI project, and wants to contribute data. They go to the project site, and click “contribute.” Their data is automatically downloaded from a cloud server, and decrypted by a verifier.
- Limitation: AI only gets to see the data contributed by people who believe in the project.
- An alternative: Sites like librivox.org and github.com recommend, as a default, that all data is explicitly released into the public domain, so it can be used by any future AI training project.

Summary

- Differential privacy

$$\Pr(Y = \text{coffee} | X = \text{coffee}) = \frac{pq}{pq + r(1-q)}$$

- Homomorphic encryption

$$\varepsilon(x_1 + x_2) = \varepsilon(x_1) + \varepsilon(x_2)$$
$$\varepsilon(x_1 x_2) = \varepsilon(x_1)\varepsilon(x_2)$$
$$\varepsilon(\max(0, x_1)) = \max(0, \varepsilon(x_1))$$

- Federated learning/Rights-based algorithmic framework
  - Your cloud service provider never knows your unencrypted data.
  - You send the encrypted data to Acme.
  - You grant Acme permission to use your data, for a specified use, by decrypting it in the app where Acme is using it.