



# Privacy & ML

---

David Sontag  
April 22, 2025

Acknowledgements:  
Peter Szolovits and Fei Wang



**Massachusetts  
Institute of  
Technology**

# Lecture attendance submission

- Lecture attendance submission code:

## **Annotation**

- Submit URL:



ML needs a lot of data

# Outline

---

- Privacy, Confidentiality, Security
- Implicit contract between patients and the health care system
- De-identification or Anonymization of Data
- Differential Privacy / Learning
- Federated Learning from non-shared Data
- Can Models Leak?



# Protecting...

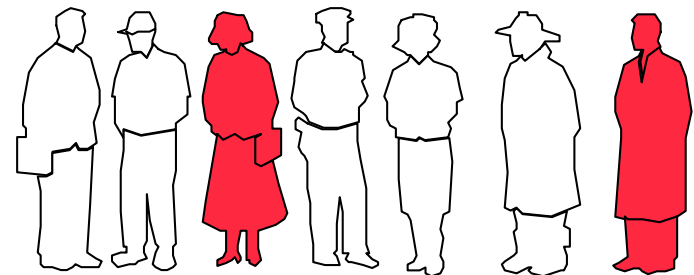
---

- What?
  - Privacy
    - Individual's desire to limit disclosure of personal information
    - What about groups?
  - Confidentiality
    - Information sharing in a controlled manner
  - Security
    - Protecting information against accident, disaster, theft, alteration, sabotage, denial of service, ...
- Against what?
  - "Evil hackers"
  - Malicious insiders
  - Stupidity
  - Subpoenas
  - Information Warfare

# Privacy

---

- Right to be let alone; e.g.:
  - snooping on Dan Quayle by J. Rothfeder (1999)
  - “outing” of Arthur Ashe (HIV), Rep. Henry Hyde (adultery), Rep. Ed Schrock (used a gay dating service)
  - celebrity medical problems (Tammy Wynette, Nicole Simpson)
- ... applies mostly to known individuals
- “No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence ... Everyone has the right to the protection of the law against such interference.”
  - ♦ Article 12, Universal Declaration of Human Rights
- “Privacy is dead, deal with it,”
  - Scott McNealy (Dec. 2000)
- Privacy in Obscurity?
  - But, Correlation among pervasive databases:
    - census, marketing, health



# Do People Care About Privacy?

## Passwords revealed by sweet deal

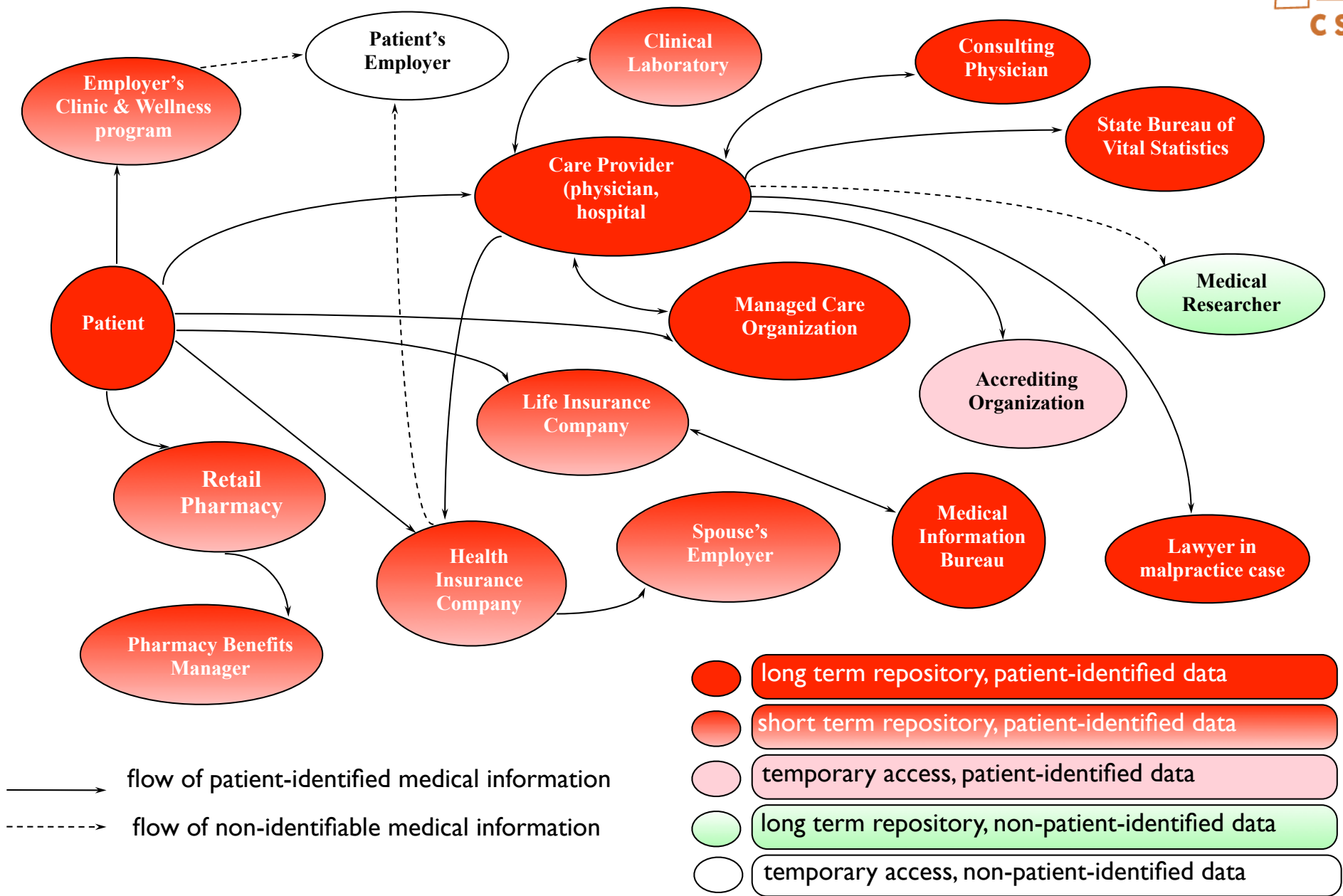
More than 70% of people would reveal their computer password in exchange for a bar of chocolate, a survey has found.



# HIPAA is not about Privacy

---

- The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is a federal law that required the creation of national standards to protect sensitive patient health information from being disclosed without the patient's consent or knowledge.
- “The consent provisions...are replaced with a new provision... that provides regulatory permission for covered entities to use and disclose protected health information for **treatment, payment, and health care operations.**” 67 Fed. Reg.



# UnitedHealthcare CEO says 'maybe a third' of US citizens were affected by recent hack

Lorenzo Franceschi-Bicchierai @lorenzofb / 3:38 PM EDT • May 1, 2024

 Comment



 Image Credits: Al Drago/Bloomberg / Getty Images

# De-Identification (and Anonymization)

---

- “**De-Identification**” = remove all explicit identifiers
- By HIPAA regulations: name, address, phone number, fax number, email address, URL, IP address, social security number, medical record number, health plan number, account number, certificate/license number, vehicle id, device id, biometric id, full-face photo, date of birth, zip code, gender, race, profession
  - “any other unique identifying number, characteristic, or code”
  - “actual knowledge that the information could be used ... to identify”
- But, patterns of doctor visits, immunizations, etc. make patients identifiable by inference, depending on knowledge and abilities of data user
- Small bin sizes lead to identifiability
  - Aggregate data into larger bins
    - dob => age
    - 3 digits of zip code
- Limited Data Set: allows inclusion of dates, full zip codes, but requires limited data use agreements



# Sweeney's Cambridge

- 1997 Cambridge, MA voting list on 54,805 voters
  - Name, address, ZIP, birth date, gender, ...
- Combinations that uniquely identify:
  - Birth date (mm/dd/yy) 12%
  - BD + gender 29%
  - BD + 5-digit ZIP 69%
  - BD + 9-digit ZIP 97%

- Unique individuals
  - Kid in a retirement community
- Rare individuals

ZIP Code	Birth Date	Gender	Race
33171	7/15/71	m	Caucasian
02657	2/18/73	f	Black
20612	3/12/75	m	Asian

Table 2. Deidentified Data that Are Not Anonymous.

- Black woman resident in Provincetown

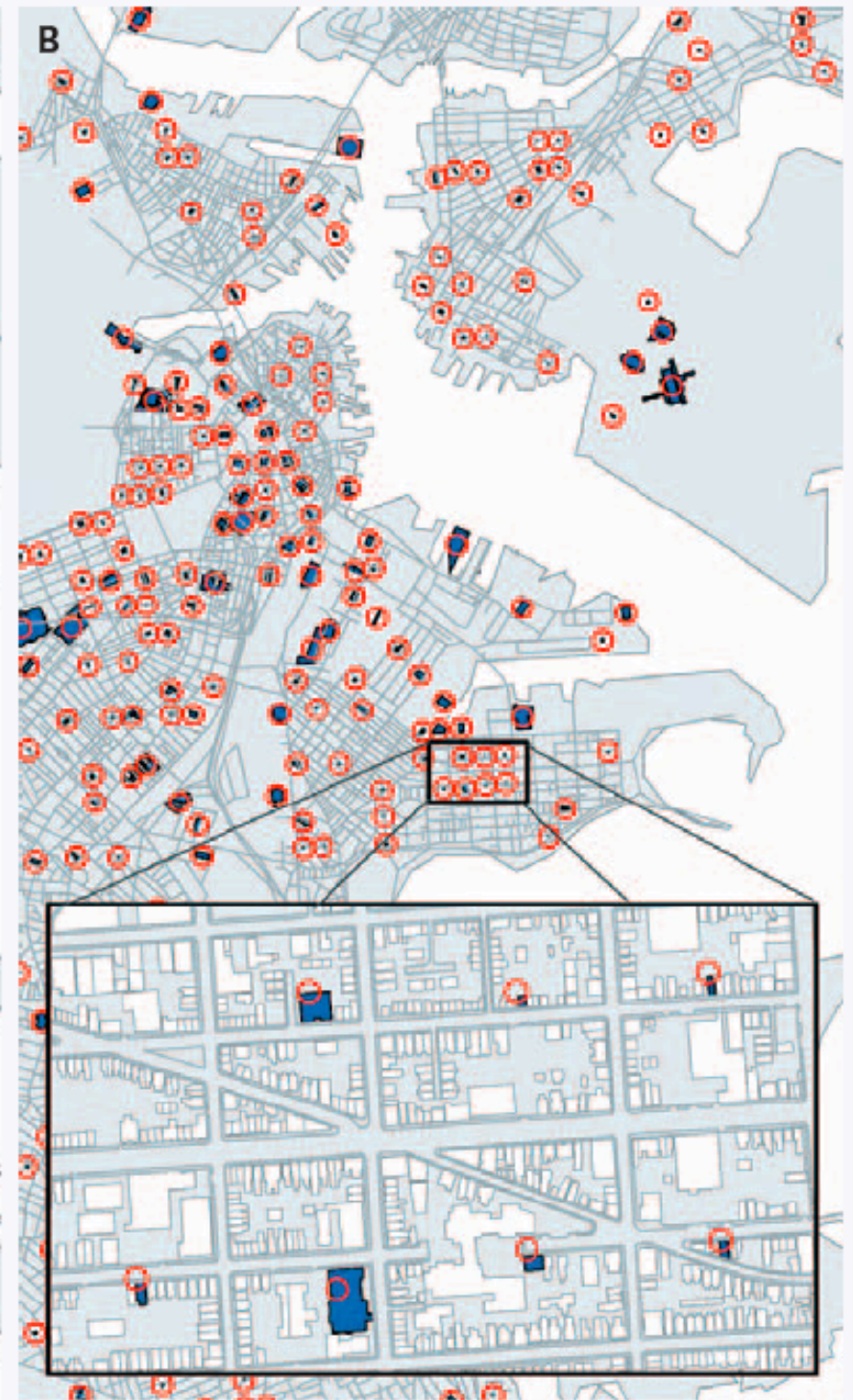


94043

Male

11/29/1976

87% of the population is uniquely identified  
[Sweeney, CMU, 2000-2001]



# What else can we do, instead of releasing original data?

---

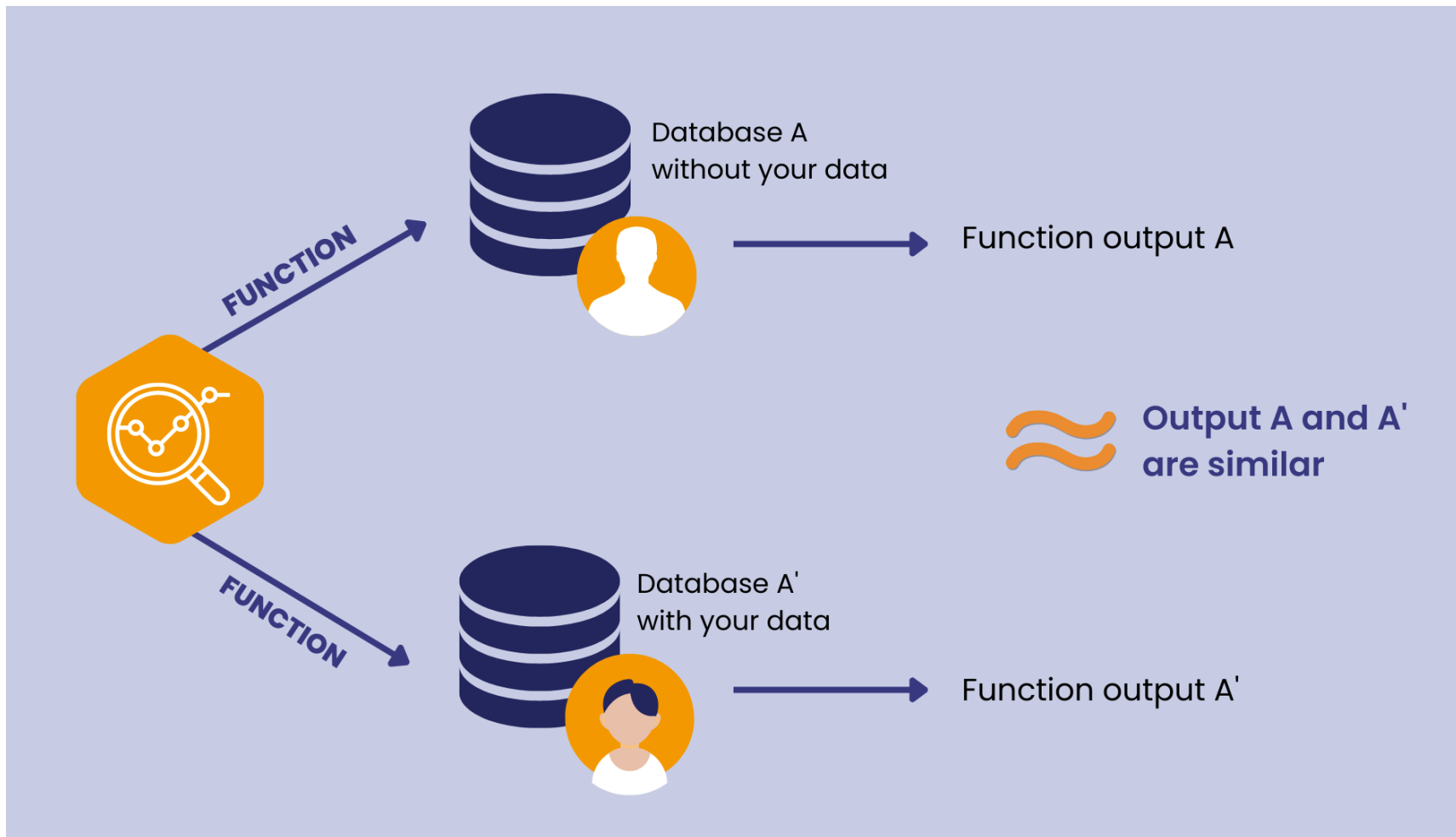
- Make sure data cannot be traced back to a set of size  $< n$ 
  - Generalization
  - Suppression of unique combinations
  - Account for leakage from what has been suppressed; e.g., back-calculating from aggregate statistics
    - E.g., dataset from International Warfarin Pharmacogenetics Consortium
      - Linear regression to predict initial dose outperforms standard clinical regimen
      - But... when one knows a target patient's background and stable dosage, their genetic markers could be predicted 22% more accurately than guessing based on marginal distributions
      - Attack is called *model inversion*
- How to estimate “external information”?
- **Every** release becomes more external info.

# Methods of Generalization/Suppression

---

- Underlying problem (find minimal generalization/suppression to achieve a level of anonymity) is NP-hard (Vinterbo)
- Mainly heuristic search over space of possible generalizations/suppressions
  - Scrub, Datafly,  $\mu$ -Argus (Netherlands), k-Similar
- T. Lasko '11: spectral anonymization
  - Build a model of data that captures the n-th order statistics of the distribution
  - **Synthesize “fake” patients from that distribution**
  - But, how do we know it doesn't just reproduce the original training data?
- Differential Privacy
- Practical approaches:
  - Put data in a secure data enclave for R&D... release ML models derived from the data
  - But, might these models reveal too much about the training data?

# Differential Privacy



- Used by Census Bureau, Apple, Google, etc.

# Differential Privacy

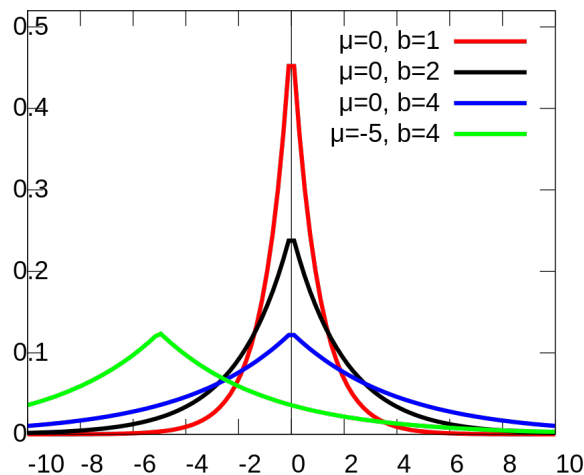
---

- An algorithm is **differentially private** if its output is statistically indistinguishable when applied to two input datasets that differ by only one record in the dataset, where  $S \subset \text{Range}(\mathcal{A})$ 
  - $\Pr[\mathcal{A}(D_1) \in S] \leq \exp(\epsilon) \cdot \Pr[\mathcal{A}(D_2) \in S] + \delta$
  - $\mathcal{A}$  is a randomized algorithm that processes the data to create an *anonymized* version: de-identification, perturbation, subsampling, ...
  - $D_1$  and  $D_2$  are subsets of the data that differ only by one element
  - $\epsilon$  and  $\delta$  are (small) numbers;  $\delta$  is prob. that privacy guarantee fails
- This condition can be achieved for all pairs of  $D_1, D_2$  by having  $\mathcal{A}$  add (Laplacian) noise to answers, depending on sensitivity to specifics of the query and the case that differs between  $D_1$  and  $D_2$ 
  - The amount of noise also depends on  $\epsilon$
  - “Privacy Budget”

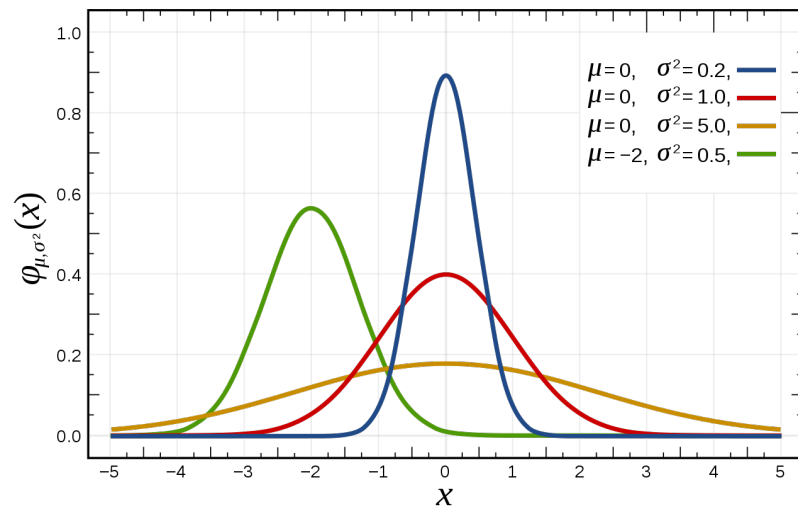
# Laplace Distribution

- $f(x | \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right)$
- Like the normal distribution, but with longer “tails”

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right)$$



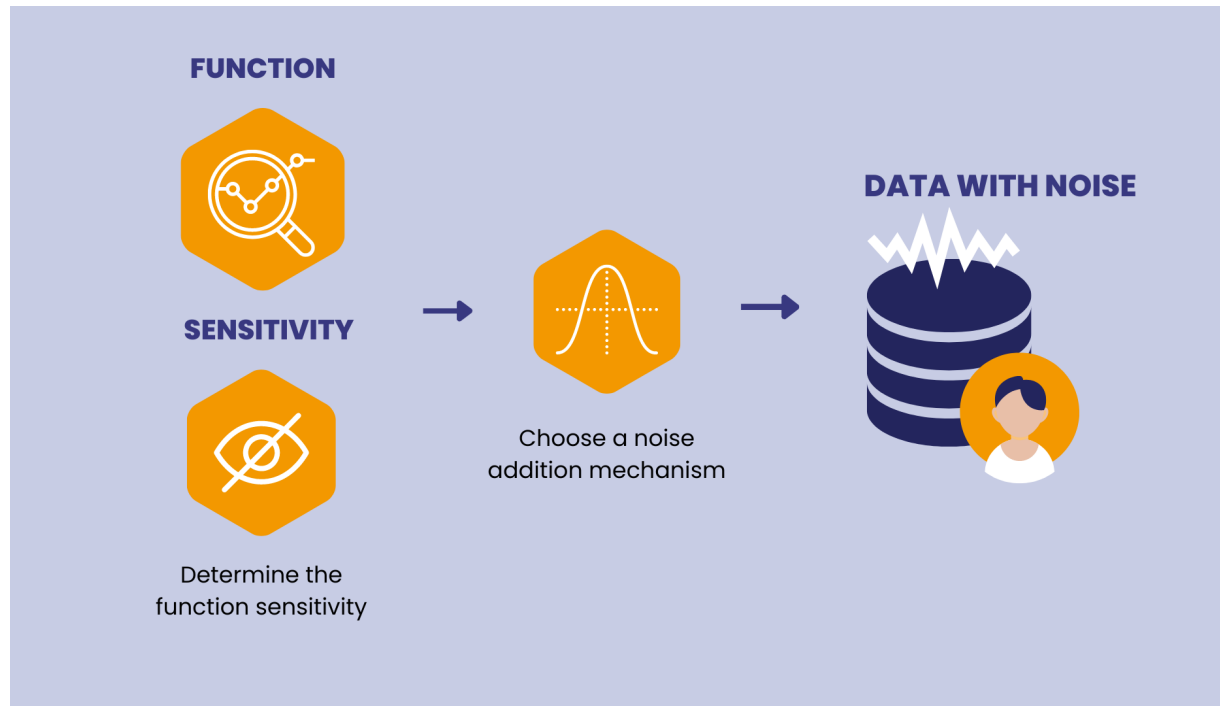
Laplace



Normal



# Sensitivity of a Query Determines Amount of Noise



- Sensitivity = maximum change that can occur in the output if a single person is added to or removed from any possible input dataset
- Therefore, DP tends to “wash out” the distribution tails
  - These may be important for useful models



# Differentially private machine learning

---

- One way to achieve DP in neural network models is via differentially private stochastic gradient descent (DP-SGD):

---

**Algorithm 1** Differentially private SGD (Outline)

---

**Input:** Examples  $\{x_1, \dots, x_N\}$ , loss function  $\mathcal{L}(\theta) = \frac{1}{N} \sum_i \mathcal{L}(\theta, x_i)$ . Parameters: learning rate  $\eta_t$ , noise scale  $\sigma$ , group size  $L$ , gradient norm bound  $C$ .

**Initialize**  $\theta_0$  randomly

**for**  $t \in [T]$  **do**

    Take a random sample  $L_t$  with sampling probability  $L/N$

**Compute gradient**

    For each  $i \in L_t$ , compute  $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t} \mathcal{L}(\theta_t, x_i)$

**Clip gradient**

$\bar{\mathbf{g}}_t(x_i) \leftarrow \mathbf{g}_t(x_i) / \max(1, \frac{\|\mathbf{g}_t(x_i)\|_2}{C})$

**Add noise**

$\tilde{\mathbf{g}}_t \leftarrow \frac{1}{L} (\sum_i \bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}))$

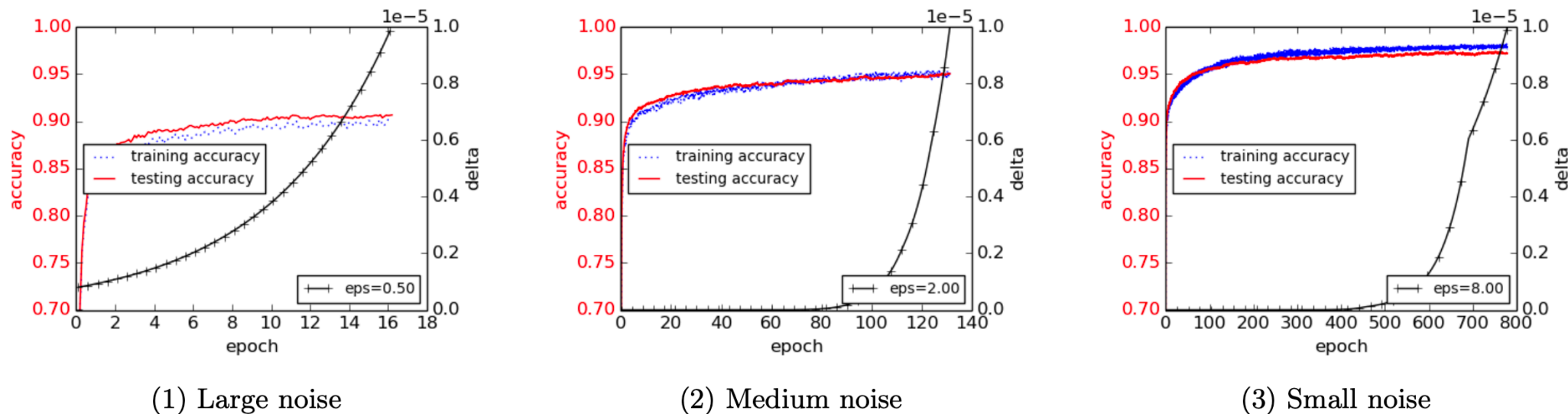
**Descent**

$\theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

**Output**  $\theta_T$  and compute the overall privacy cost  $(\epsilon, \delta)$  using a privacy accounting method.

---

# Works Well for 1-layer Model for MNIST (less well on CIFAR-10)



(1) Large noise

(2) Medium noise

(3) Small noise

**Figure 3: Results on the accuracy for different noise levels on the MNIST dataset. In all the experiments, the network uses 60 dimension PCA projection, 1,000 hidden units, and is trained using lot size 600 and clipping threshold 4. The noise levels  $(\sigma, \sigma_p)$  for training the neural network and for PCA projection are set at  $(8, 16)$ ,  $(4, 7)$ , and  $(2, 4)$ , respectively, for the three experiments.**

# The privacy-utility trade-off

- Evaluate using the following datasets:

DATASET	DATA TYPE	OUTCOME VARIABLE	$n$	$d$	CLASSIFICATION TASK	TAIL SIZE	PROTECTED ATTRIBUTES	EVALUATION
<b>HEALTH CARE</b>								
mimic_mortality	TIME SERIES	IN-ICU MORTALITY	21,877	(24,69)	BINARY	LARGE	ETHNICITY	U,R, F
mimic_los_3	TIME SERIES	LENGTH OF STAY > 3 DAYS	21,877	(24,69)	BINARY	SMALL	ETHNICITY	U,R, F
mimic_intervention	TIME SERIES	VASOPRESSOR ADMINISTRATION	21,877	(24,69)	MULTICLASS (4)	SMALL	ETHNICITY	U,R, F
NIH_chest_x_ray	IMAGING	MULTILABEL DISEASE PREDICTION	112,120	(256,256)	MULTICLASS MULTILABEL (14)	LARGEST	SEX	U,F
<b>VISION BASELINES</b>								
mnist	IMAGING	NUMBER CLASSIFICATION	60,000	(28,28)	MULTICLASS (10)	NONE	N/A	U
fashion_mnist	IMAGING	CLOTHING CLASSIFICATION	60,000	(28,28)	MULTICLASS (10)	NONE	N/A	U

**Table 1: We analyze tradeoffs in two vision baseline datasets and two health care datasets. We use three prediction tasks in MIMIC-III with different tail sizes and focus our utility (U), robustness (R), and fairness (F) analyses on these tasks. Finally, we choose NIH Chest X-Ray which is a larger dataset with the largest tail to examine whether increasing the dataset size has an impact on utility and fairness tradeoffs.**

# The privacy-utility trade-off

## VISION BASELINES

DATASET	MODEL	NONE ( $\epsilon, \delta$ )	Low ( $\epsilon, \delta$ )	HIGH ( $\epsilon, \delta$ )
MNIST	CNN	98.83 $\pm$ 0.06 ( $\infty, 0$ )	98.58 $\pm$ 0.06 ( $2.6 \cdot 10^5$ )	93.78 $\pm$ 0.25 (2.01)
FASHIONMNIST	CNN	87.92 $\pm$ 0.19 ( $\infty, 0$ )	87.90 $\pm$ 0.16 ( $2.6 \cdot 10^3$ )	79.53 $\pm$ 0.10 (2.01)

## MIMIC-III

TASK	MODEL	NONE ( $\epsilon, \delta$ )	Low ( $\epsilon, \delta$ )	HIGH ( $\epsilon, \delta$ )
MORTALITY	LR	0.82 $\pm$ 0.03 ( $\infty, 0$ )	0.76 $\pm$ 0.05 ( $3.50 \cdot 10^5, 10^{-5}$ )	0.60 $\pm$ 0.04 ( $3.54, 10^{-5}$ )
	GRUD	0.79 $\pm$ 0.03 ( $\infty, 0$ )	0.59 $\pm$ 0.09 ( $1.59 \cdot 10^3, 10^{-3}$ )	0.53 $\pm$ 0.03 ( $2.65, 10^{-3}$ )
LENGTH OF STAY > 3	LR	0.69 $\pm$ 0.02 ( $\infty, 0$ )	0.66 $\pm$ 0.03 ( $3.50 \cdot 10^5, 10^{-5}$ )	0.60 $\pm$ 0.04 ( $3.54, 10^{-5}$ )
	GRUD	0.67 $\pm$ 0.03 ( $\infty, 0$ )	0.63 $\pm$ 0.02 ( $1.59 \cdot 10^5, 10^{-5}$ )	0.61 $\pm$ 0.03 ( $2.65, 10^{-5}$ )
INTERVENTION ONSET (VASO)	LR	0.90 $\pm$ 0.03 ( $\infty, 0$ )	0.87 $\pm$ 0.03 ( $1.63 \cdot 10^7, 10^{-5}$ )	0.77 $\pm$ 0.05 ( $0.94, 10^{-5}$ )
	CNN	0.88 $\pm$ 0.04 ( $\infty, 0$ )	0.86 $\pm$ 0.02 ( $5.95 \cdot 10^7, 10^{-3}$ )	0.68 $\pm$ 0.04 ( $0.66, 10^{-3}$ )

## NIH CHEST X-RAY

METRIC	MODEL	NONE ( $\epsilon, \delta$ )	Low ( $\epsilon, \delta$ )	HIGH ( $\epsilon, \delta$ )
AVERAGE AUC	DENSENET-121	0.84 $\pm$ 0.00 ( $\infty, 0$ )	0.51 $\pm$ 0.01 ( $1.74 \cdot 10^5, 10^{-6}$ )	0.49 $\pm$ 0.00 ( $0.84, 10^{-6}$ )
BEST AUC	DENSENET-121	0.98 $\pm$ 0.00 (HERNIA)	0.54 $\pm$ 0.04 (EDEMA)	0.54 $\pm$ 0.05 (PLEURAL THICKENING)
WORST AUC	DENSENET-121	0.72 $\pm$ 0.00 (INFILTRATION)	0.48 $\pm$ 0.02 (FIBROSIS)	0.47 $\pm$ 0.02 (PLEURAL THICKENING)

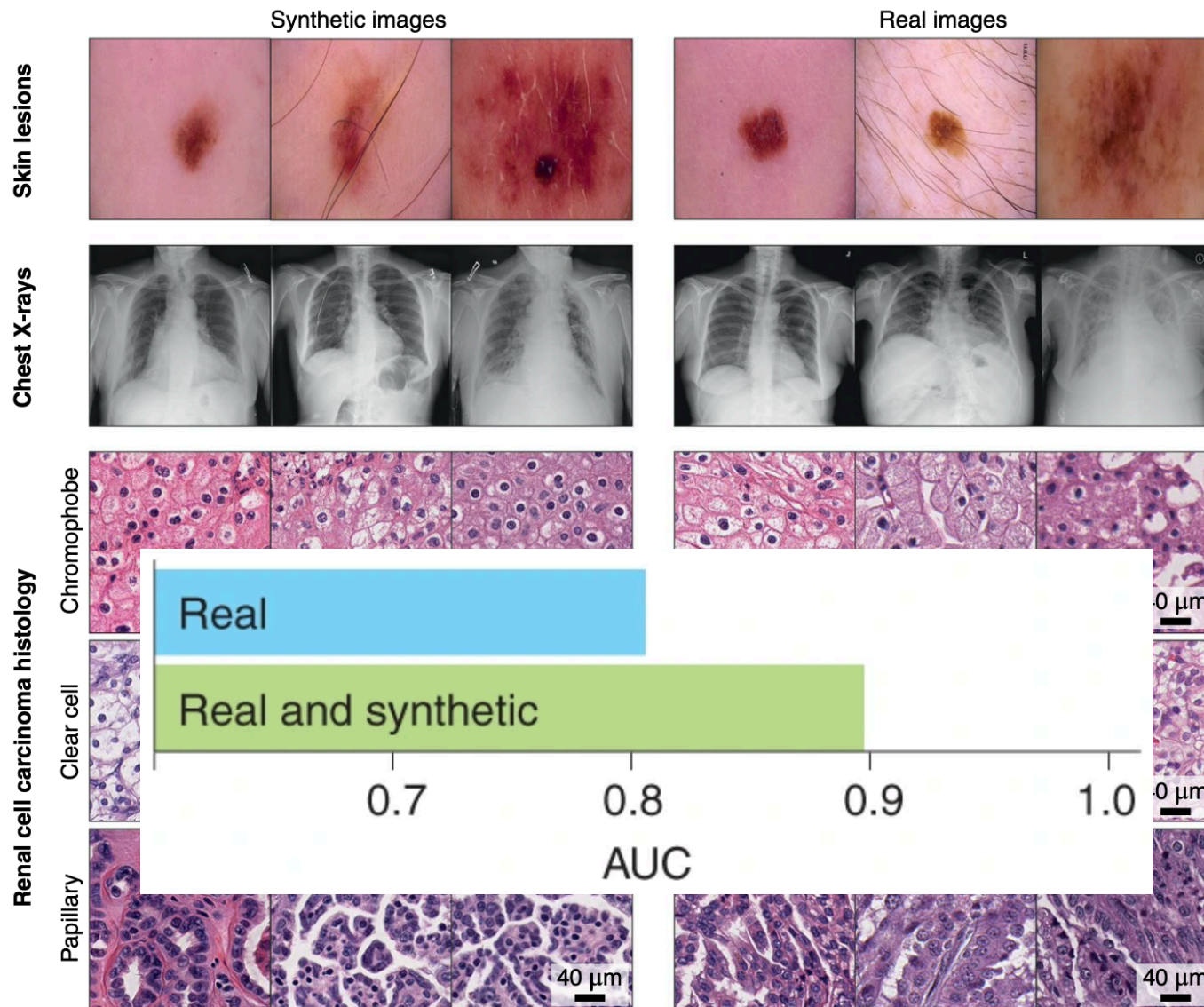
**Table 2: Health care tasks have a significant tradeoff between the High and Low or None setting. The tradeoff is better in tasks with small tails (length of stay and intervention onset), and worst in tasks such as mortality and NIH Chest X-Ray with long tails. We provide the  $\epsilon, \delta$  guarantees in parentheses, where  $\epsilon$  represents the privacy loss (lower is better) and  $\delta$  represents the probability that the guarantee does not hold (lower is better).**

# Thought on Differential Privacy in Healthcare

---

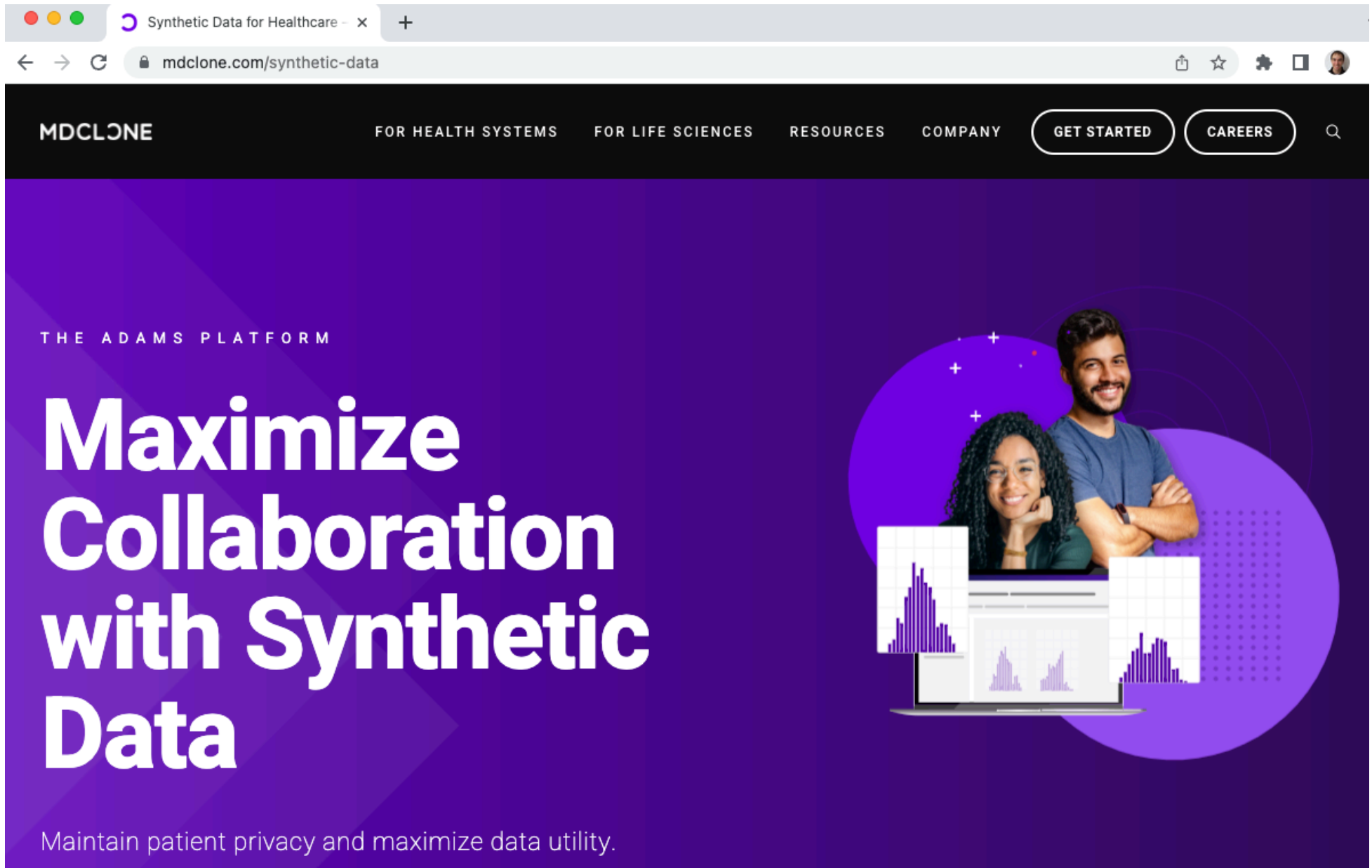
- Great to quantify privacy risks
  - though setting and interpretation of  $\epsilon$  remains problematic
- Very high cost in model performance
- DP can serve as a regularizer, thought to improve performance with dataset shift
  - Evidence from biomedical fields does not support this hope
- By washing out tails, DP focuses attention on largest groups, reducing fairness

# Synthetic Data Generation by GANs





# Synthetic data generation



The image is a screenshot of a web browser displaying the MDClone website. The browser's address bar shows the URL "mdclone.com/synthetic-data". The website's navigation bar includes the MDClone logo and links for "FOR HEALTH SYSTEMS", "FOR LIFE SCIENCES", "RESOURCES", and "COMPANY". There are also buttons for "GET STARTED" and "CAREERS", along with a search icon. The main content area features a purple background with the text "THE ADAMS PLATFORM" and a large headline: "Maximize Collaboration with Synthetic Data". Below the headline, it says "Maintain patient privacy and maximize data utility." On the right side, there is an illustration of a man and a woman smiling, with a laptop in front of them displaying data charts. The background of the illustration includes a grid pattern and several plus signs.

Synthetic Data for Healthcare - x +

mdclone.com/synthetic-data

MDCLONE

FOR HEALTH SYSTEMS FOR LIFE SCIENCES RESOURCES COMPANY

GET STARTED CAREERS

THE ADAMS PLATFORM

# Maximize Collaboration with Synthetic Data

Maintain patient privacy and maximize data utility.

# Classifiers can reveal information about training data

- An attack called *model inversion* can be used to reverse engineer training data
- Similar problem with synthetic data

---

**Algorithm 1** Inversion attack for facial recognition models.

---

```

1: function MI-FACE(label,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\lambda$ )
2:    $c(\mathbf{x}) \stackrel{\text{def}}{=} 1 - \tilde{f}_{\text{label}}(\mathbf{x})$ 
3:    $\mathbf{x}_0 \leftarrow \mathbf{0}$ 
4:   for  $i \leftarrow 1 \dots \alpha$  do
5:      $\mathbf{x}_i \leftarrow \text{PROCESS}(\mathbf{x}_{i-1} - \lambda \cdot \nabla c(\mathbf{x}_{i-1}))$ 
6:     if  $c(\mathbf{x}_i) \geq \max(c(\mathbf{x}_{i-1}), \dots, c(\mathbf{x}_{i-\beta}))$  then
7:       break
8:     if  $c(\mathbf{x}_i) \leq \gamma$  then
9:       break
10:  return  $[\arg \min_{\mathbf{x}_i} (c(\mathbf{x}_i)), \min_{\mathbf{x}_i} (c(\mathbf{x}_i))]$ 

```

---

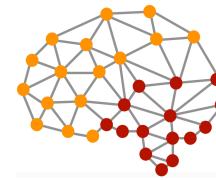


Figure 7: Reconstruction without using Process-DAE (Algorithm 2) (left), with it (center), and the training set image (right).





**Weill Cornell  
Medicine**



**WangLab**  
Health Data Analytics  
Weill Cornell Medicine

# Federated Learning in Large Clinical Research Networks

Fei Wang

Associate Professor

Department of Population Health Sciences

Weill Cornell Medicine

[feiwang.cornell@gmail.com](mailto:feiwang.cornell@gmail.com)

 @feiwang03



<https://wcm-wanglab.github.io/index.html>

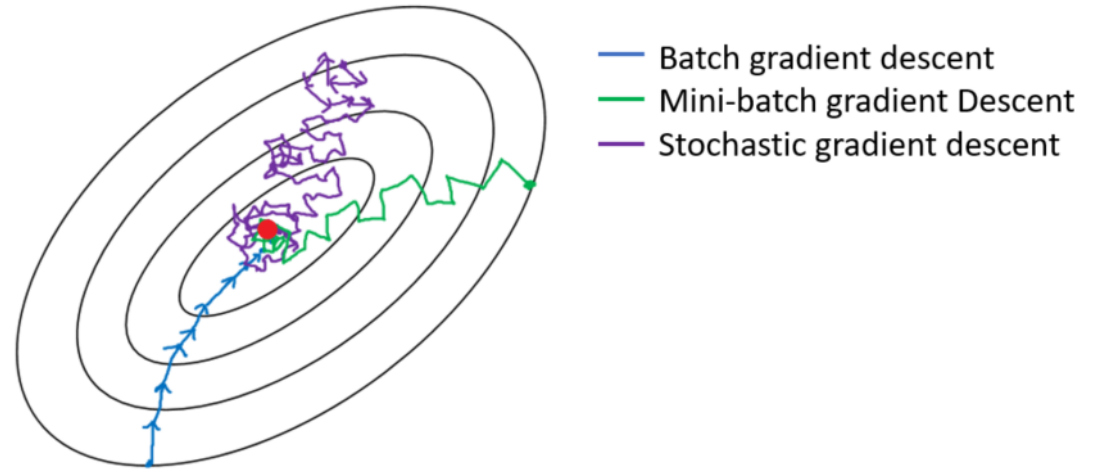
29

Material on Federated Learning from Prof. Fei Wang  
(with permission)

# Stochastic Gradient Descent

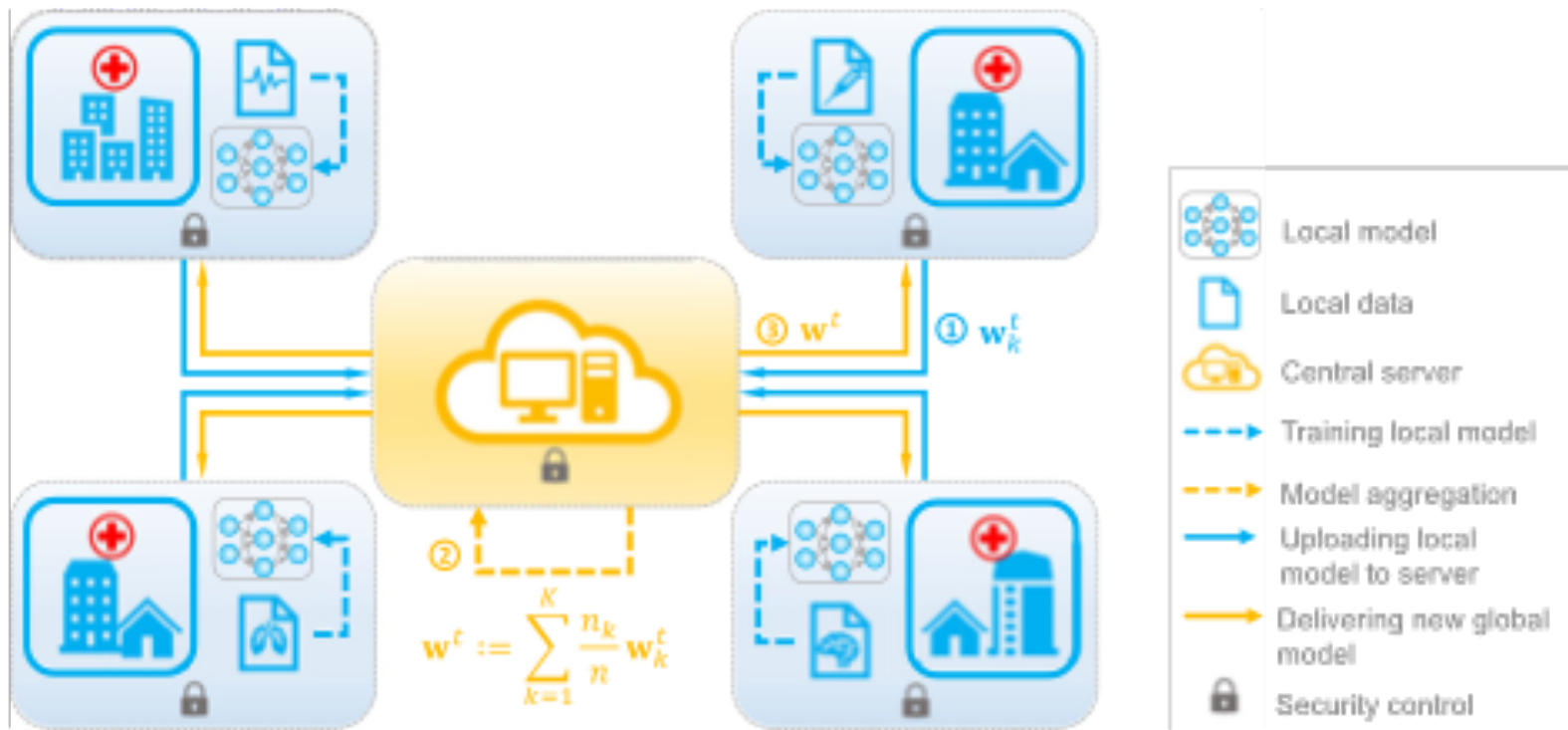
- At each step of gradient descent, instead of compute for all training samples, randomly pick a small subset (mini-batch) of training samples

$$w_{t+1} \leftarrow w_t - \eta \nabla f(w_t; x_k, y_k)$$



<https://medium.com/analytics-vidhya/gradient-descent-vs-stochastic-gd-vs-mini-batch-sgd-fbd3a2cb4ba4>

# Federated Learning

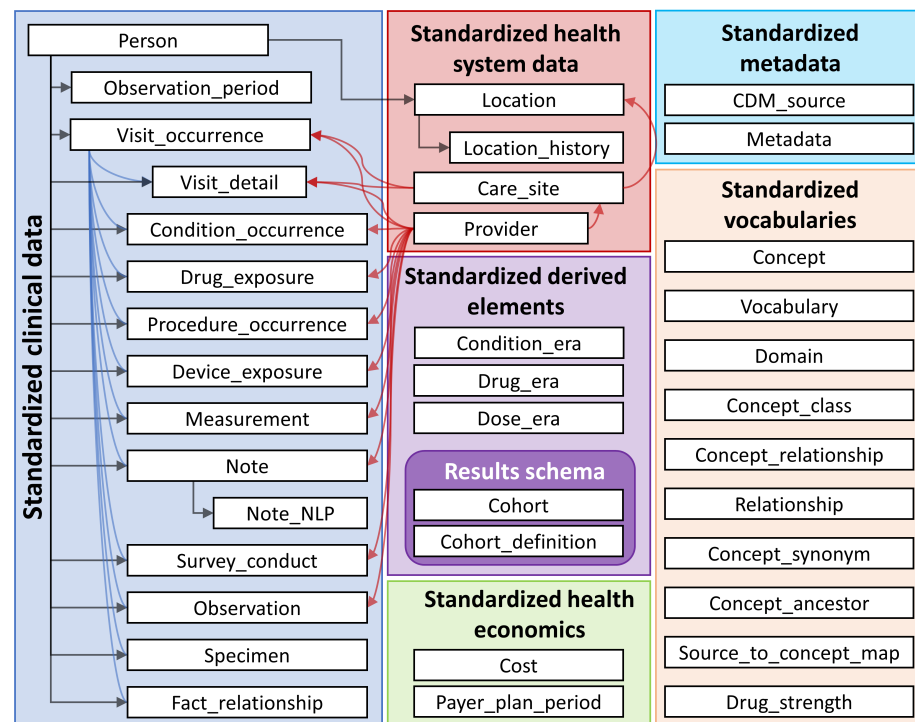


Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and **Fei Wang**. "Federated learning for healthcare informatics." *Journal of Healthcare Informatics Research* (2020): 1-19.

# Clinical Research Networks



<https://ohdsi.github.io/TheBookOfOhdsi/OhdsiCommunity.html>



<https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html>

# Federated SGD

- In a round  $t$ :
  - The central server broadcasts current model  $w_t$  to each client; each client  $k$  computes gradient:  $g_k = \nabla F_k(w_t)$ , on its local data.
    - Approach 1: Each client  $k$  submits  $g_k$ ; the central server aggregates the gradients to generate a new model:
      - $w_{t+1} \leftarrow w_t - \eta \nabla f(w_t) = w_t - \eta \sum_{k=1}^K \frac{n_k}{n} g_k$ .
    - Approach 2: Each client  $k$  computes:  $w_{t+1}^k \leftarrow w_t - \eta g_k$ ; the central server performs aggregation:
      - $w_{t+1} \leftarrow \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^k$

# Federated Averaging

---

**Algorithm 1** FederatedAveraging. The  $K$  clients are indexed by  $k$ ;  $B$  is the local minibatch size,  $E$  is the number of local epochs, and  $\eta$  is the learning rate.

---

**Server executes:**

initialize  $w_0$

**for** each round  $t = 1, 2, \dots$  **do**

$m \leftarrow \max(C \cdot K, 1)$

$S_t \leftarrow$  (random set of  $m$  clients)

**for** each client  $k \in S_t$  **in parallel do**

$w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$

$w_{t+1} \leftarrow \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^k$

**ClientUpdate**( $k, w$ ): // Run on client  $k$

$\mathcal{B} \leftarrow$  (split  $\mathcal{P}_k$  into batches of size  $B$ )

**for** each local epoch  $i$  from 1 to  $E$  **do**

**for** batch  $b \in \mathcal{B}$  **do**

$w \leftarrow w - \eta \nabla \ell(w; b)$

    return  $w$  to server

---

1. At first, a model is randomly initialized on the central server
2. For each round  $t$ :
  - A. A random set of clients is chosen
  - B. Each client performs local gradient descent steps
  - C. The server aggregates model parameters submitted by the clients

# What Do We Know about Convergence of Federated Learning?

---

- Speed-accuracy conflict
  - linear convergence but to a (possibly) incorrect point
  - convergence to the global minimum but at a sub-linear rate
- “when the clients' local loss functions are smooth and strongly convex, we show that FedLin guarantees linear convergence to the global minimum, despite arbitrary objective and systems heterogeneity”
  - Mitra A, Jaafar R, Pappas GJ, Hassani H. Linear Convergence in Federated Learning: Tackling Client Heterogeneity and Sparse Gradients. In 2022. Available from: <https://openreview.net/forum?id=h7FqQ6hCK18>

## Study Population

Adults hospitalized with laboratory-confirmed COVID-19



## Study Locations

5 hospitals in New York City



## Primary Outcome

Mortality within 7 days of admission



## Models

### Local

Local data from each hospital individually trained



### Pooled

All individual hospital data aggregated for training



### Federated

Central aggregator with only model parameters shared between hospitals



## Classifiers



### LASSO

(Least absolute shrinkage and selection operator)



### MLP

(Multilayer perceptron)

## Learning Framework Comparisons

Model performance across 5 hospitals:  
AUC-ROC\* (95% CI) values

	LASSO	MLP
Local	0.666 (0.662-0.671)	0.766 (0.763-0.769)
Pooled	0.792 (0.790-0.794)	0.798 (0.796-0.800)
Federated	<b>0.766</b> (0.763-0.768)	<b>0.810</b> (0.808-0.812)

\*Area under the receiver operating characteristic curve

**Summary:** Federated model classifiers outperform locally trained classifiers in predicting mortality among hospitalized patients with COVID-19.

Akhil Vaid, Suraj K Jaladanki, Jie Xu, Shelly Teng, Arvind Kumar, Samuel Lee, Sulaiman Somani, Ishan Paranjpe, Jessica K De Freitas, Tingyi Wanyan, Kipp W Johnson, Mesude Bicak, Eyal Klang, Young Joon Kwon, Anthony Costa, Shan Zhao, Riccardo Miotto, Alexander W Charney, Erwin Böttinger, Zahi A Fayad, Girish N Nadkarni, **Fei Wang**, Benjamin S Glicksberg. "Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach." *JMIR medical informatics* 9, no. 1 (2021): e24207.



# Can Models Themselves Leak Data?

---

- Depends on model size
  - “simple probing methods are not able to meaningfully extract sensitive information from BERT trained over the MIMIC-III corpus”
    - Lehman E, Jain S, Pichotta K, Goldberg Y, Wallace BC. Does BERT Pretrained on Clinical Notes Reveal Sensitive Data? [Internet]. arXiv; 2021. Available from: <http://arxiv.org/abs/2104.07762>
- But, ask ChatGPT to act as a Jupyter notebook

```
import pandas as pd
note_events = pd.read_csv('NOTEEVENTS.csv')
print(len(note_events))
print(len(set(note_events.SUBJECT_ID)))
```

2083180  
46146
- Without access to MIMIC (we think), it seems to have memorized this information from reading notebooks on the Web

# Can Models Themselves Leak Data?

---

- GPT-4, asked about `print(df.iloc[1].TEXT)`, responds with similar (but not identical) data:
- Admission Date: [2118-4-21] Discharge Date: [2118-4-29]  
Date of Birth: [2036-9-29] **Sex: F**  
**Service: MEDICINE**  
Allergies: Patient recorded as having No Known Allergies to Drugs  
Attending:[First Name3 (LF) 209]  
Chief Complaint: Shortness of breath  
Major Surgical or Invasive Procedure: None  
History of Present Illness: **81-year-old female with a history of emphysema (not on home O2), who presents with three days of shortness of breath thought by her primary care doctor to be a COPD flare. Two days prior to admission, she was started on a prednisone taper and one dose of azithromycin.** The patient denies any fever or chills but has noticed an increased cough with some minimal yellow sputum production. She denies any chest pain, nausea or vomiting, and abdominal pain.  
Past Medical History: 1. COPD, emphysema 2. Hypertension  
Social History: The patient is a former smoker with a 40 pack-year smoking history but quit 20 years ago. She denies alcohol or drug use.
- Bad!
  - Model clearly memorized part of a case.
  - Someone violated their DUA

# The New York Times sues OpenAI and Microsoft for using its stories to train chatbots



EXAMPLE 1: HOW THE U.S. LOST OUT ON IPHONE WORK

<https://www.nytimes.com/2012/01/22/business/apple-america-and-a-squeezed-middle-class.html>  
 Copyright number: TX 6-789-918 Effective Date: 2012-05-31

**Prompt (taken from article):**

When Barack Obama joined Silicon Valley’s top luminaries for dinner in California last February, each guest was asked to come with a question for the president.

But as Steven P. Jobs of Apple spoke, President Obama interrupted with an inquiry of his own: what would it take to make iPhones in the United States?

Not long ago, Apple boasted that its products were made in America. Today, few are. Almost all of the 70 million iPhones, 30 million iPads and 59 million other products Apple sold last year were manufactured overseas.

Why can’t that work come home? Mr. Obama asked.

Mr. Jobs’s reply was unambiguous. “Those jobs aren’t coming back,” he said, according to another dinner guest.

The president’s question touched upon a central conviction at Apple. It isn’t just that workers are cheaper abroad. Rather, Apple’s executives believe the vast scale of overseas factories as well as the flexibility, diligence and industrial skills of foreign workers have so outpaced their American counterparts that “Made in the U.S.A.” is no longer a viable option for most Apple products.

Apple has become one of the best-known, most admired and most imitated companies on earth, in part through an unrelenting mastery of global operations. Last year, it earned over \$400,000 in profit per employee, more than Goldman Sachs, Exxon Mobil or Google.

However, what has vexed Mr. Obama as well as economists and policy makers is that Apple — and

**Output from GPT-4:**

many of America’s other global companies — aren’t nearly as avid in creating American jobs as other famous companies were in their heydays.

Apple employs 43,000 people in the United States and 20,000 overseas, a small fraction of the over 400,000 American workers at General Motors in the 1950s, or the hundreds of thousands at General Electric in the 1980s. Many more people work for Apple’s contractors: an additional 700,000 people engineer, build and assemble iPads, iPhones and Apple’s other products. But almost none of them work in the United States. Instead, they work for foreign companies in Asia, Europe and elsewhere, at factories that almost all electronic designers rely upon

**Actual text from NYTimes:**

many of its high-technology peers — are not nearly as avid in creating American jobs as other famous companies were in their heydays.

Apple employs 43,000 people in the United States and 20,000 overseas, a small fraction of the over 400,000 American workers at General Motors in the 1950s, or the hundreds of thousands at General Electric in the 1980s. Many more people work for Apple’s contractors: an additional 700,000 people engineer, build and assemble iPads, iPhones and Apple’s other products. But almost none of them work in the United States. Instead, they work for foreign companies in Asia, Europe and elsewhere, at factories that almost all electronic designers rely upon

The New York Times’s lawsuit against OpenAI and Microsoft provides examples of ChatGPT producing text similar to the publication’s articles. THE NEW YORK TIMES

# Summary

---

- Privacy, Confidentiality, Security
- Implicit contract between patients and the health care system
- De-identification or Anonymization of Data
- Federated Learning from non-shared Data
- Can Models Leak?