Privacy and Confidentiality

Peter Szolovits
April 13, 2023

With some slides from David Sontag and Fei Wang

Massachusetts Institute of Technology

Outline

- 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?

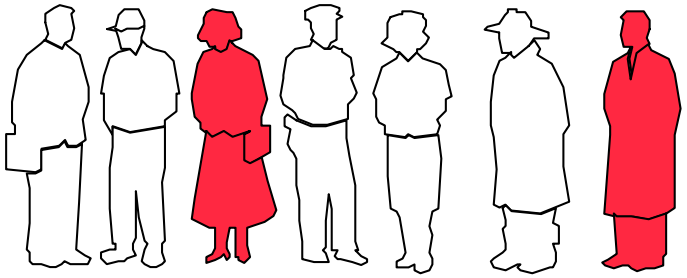
Protecting…

- 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

National Research Council. For the record: Protecting electronic health information. 1997. Available from: http://www.nap.edu/catalog/5595.html

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



People Don’t Care About Privacy

Passwords revealed by sweet deal

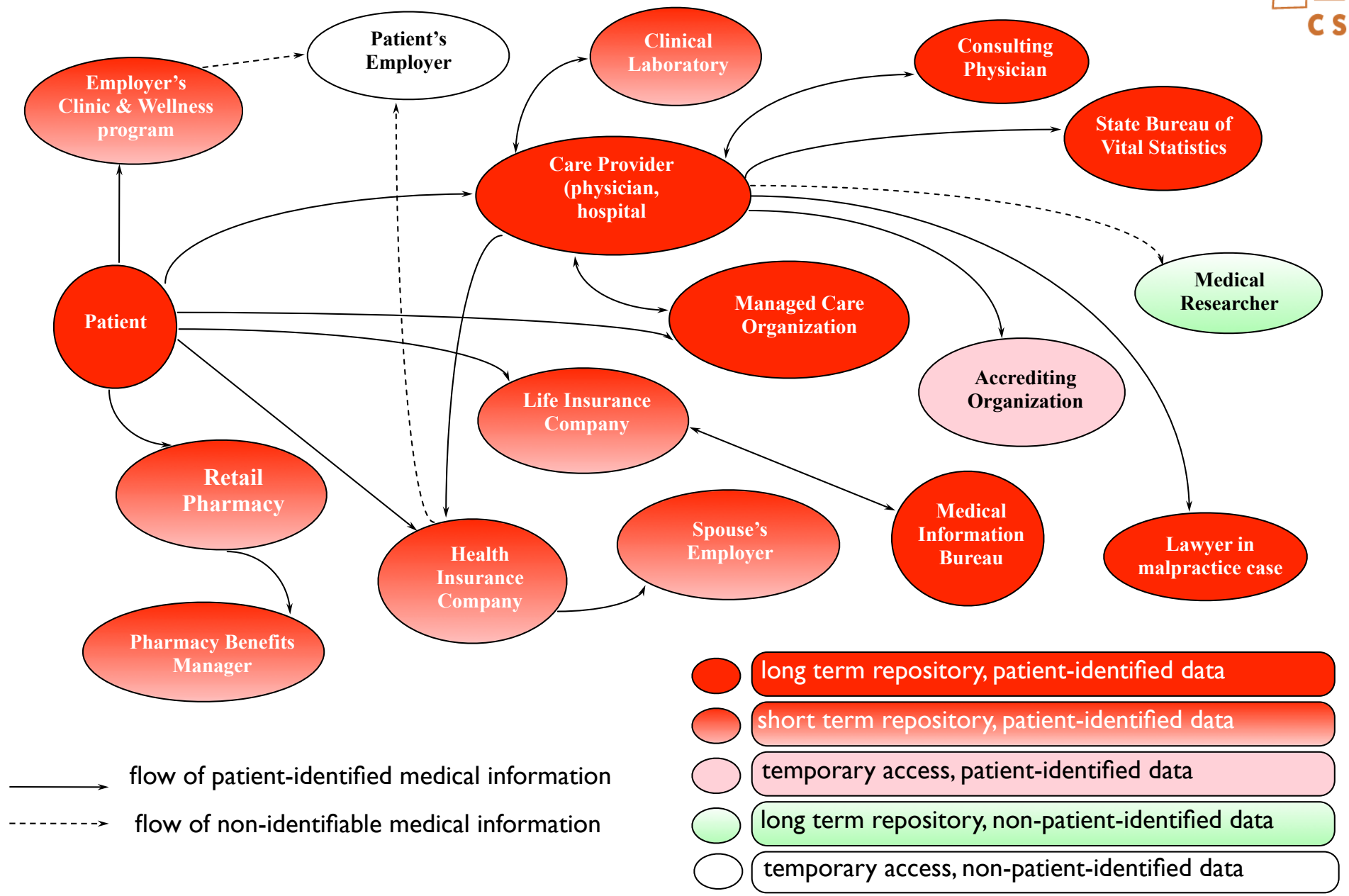
More than 70% of people would reveal their computer password in exchange for a bar of chocolate, a survey has found.

https://blog.tmb.co.uk/passwords-for-chocolate

Confidentiality

- Use and sharing of information by multiple users at many institutions
- Should be controlled by coherent policy
- Enforced by appropriate technology

- E.g., who may use results of your life insurance physical exam, for what purposes?



National Research Council. For the record: Protecting electronic health information. 1997. Available from: http://www.nap.edu/catalog/5595.html

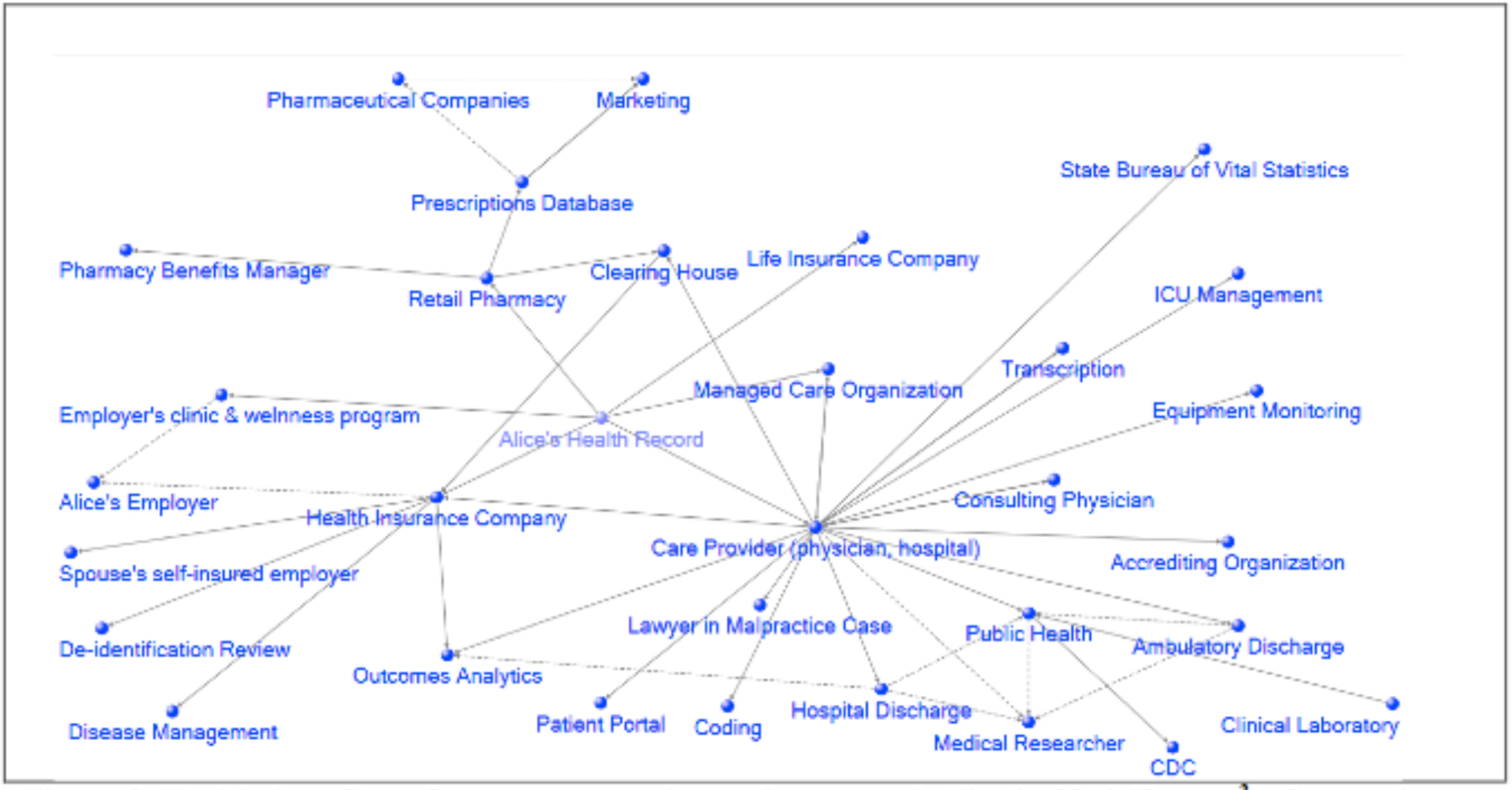


Figure 2. Health data flows for a representative patient named Alice in 2010 [Source[3]]. Comparing Figure 1 to Figure 2, the kinds of entities receiving information doubled, and today there is increased use of identifiable patient information and only long-term storage.

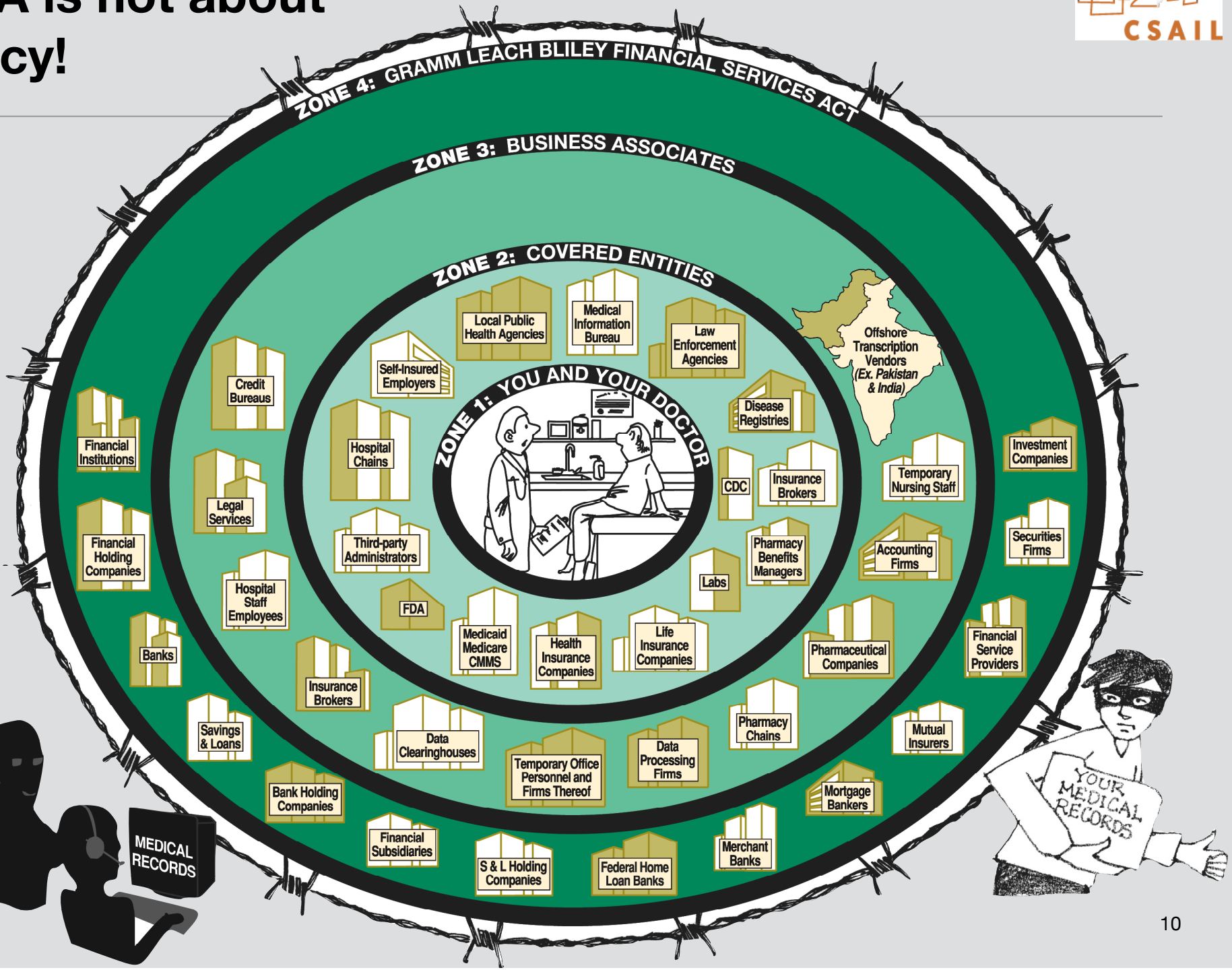
http://dataprivacylab.org/projects/irb/DataPrivacyLab.pdf

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.

HIPAA is not about Privacy!

"regulatory permission for covered entities to use and disclose protected health information for treatment, payment, and health care operations"



Security

- Integrity of data
  - No unauthorized modifications
  - No “dropped bits”
- Availability
  - Natural disaster
  - Adversary attack
  - Inadequacy of backup, fail-over
- Enforcement of confidentiality policies

Understanding Between Patients and Hospitals (e.g., MGH Authorization for Obstetrical Care)

"I have read Care During Labor and Delivery.

I understand what has been discussed with me, as well as the content of this form. I have been given the opportunity to ask questions and have received satisfactory answers.

I understand that no guarantees or promises have been made to me about expected results of this pregnancy.

I am aware that other risks and complications may occur. I also understand that during the remainder of my pregnancy or during labor, unforeseen conditions may be revealed that require additional procedures.

I know that resident doctors and other clinical students/staff may help my doctor or midwife.

I understand that tissue or other specimens removed from me as necessary during obstetrical procedures, including placental tissue, may subsequently be used by the Hospital, its affiliates, or other academic or commercial entities for research, educational purposes (including photographing), or other activity, if it furthers the Hospital's missions.

All of my questions have been answered and I consent to obstetrical care during my birthing experience. I understand that some of the procedures described above may occur. I retain the right to refuse any specific treatment. Ongoing discussion(s) about my current status and the recommended steps will be a part of my care."

https://www.massgeneral.org/assets/mgh/pdf/obgyn/l_d consent-english.pdf

Crimson Core

- The Crimson Biomaterials Collection Core Facility prospectively collects discarded clinical materials matching investigator-defined criteria against available information on clinical samples, including ICD.9 codes and results of clinical laboratory testing.
- Studies using the core must either
  - (1) have an IRB-approved protocol for discarded clinical materials and anonymized information or
  - (2) a protocol to allow collection of discarded samples from patients consented for their study.
- Collected samples may be additionally processed, aliquotted, or tested per the menu of clinical laboratory tests available within the BWH Clinical Laboratories.

- *“Available information” is matched to sample data but de-identified to investigators.*

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
  - Black woman resident in Provincetown

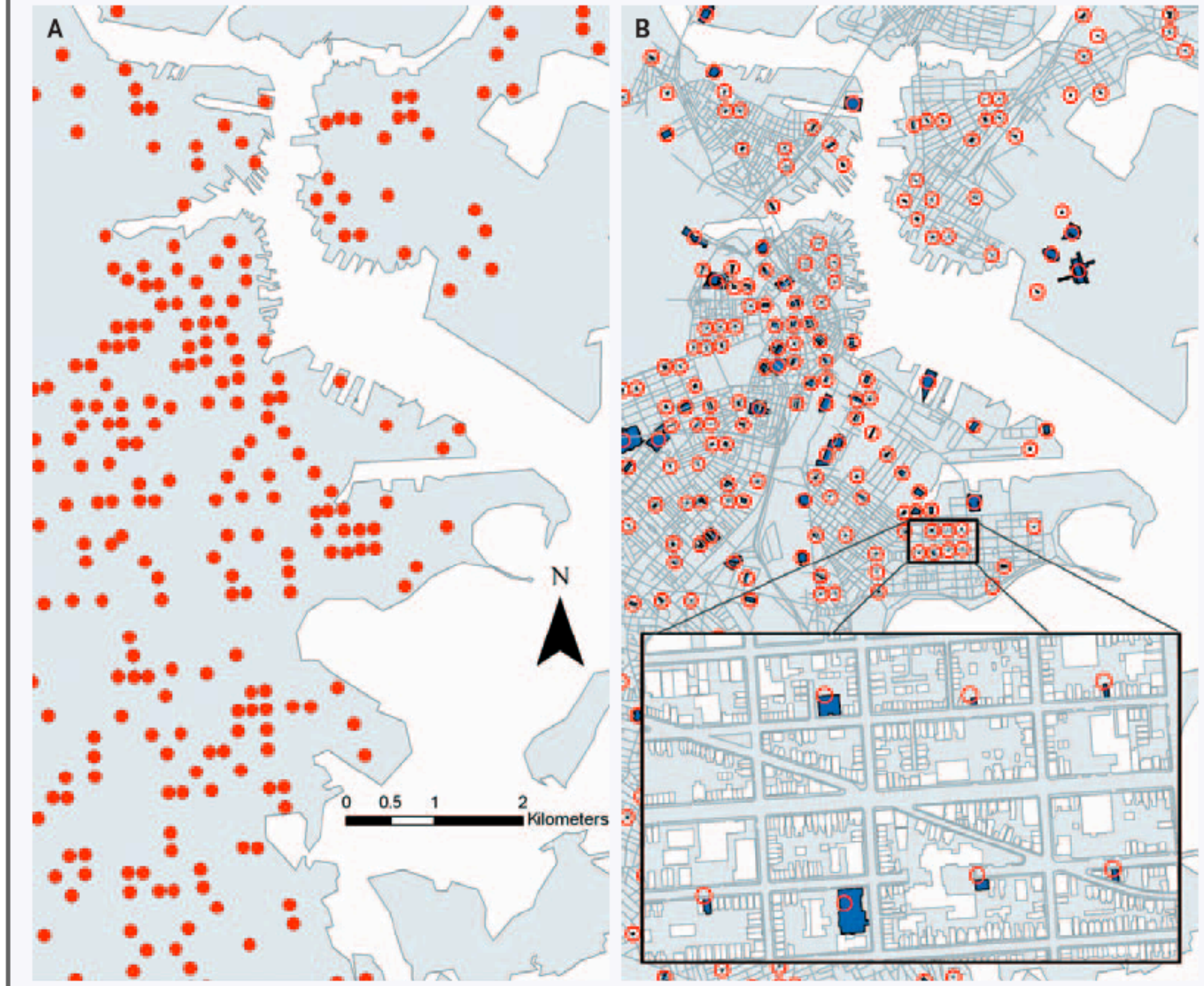
ZIP Code	*Birth Date*	*Gender*	*Race*
33171	7/15/71	m	Caucasian
02657	2/18/73	f	Black
20612	3/12/75	m	Asian

Table 2. Deidentified Data that Are Not Anonymous.

L. Sweeney. *Maintaining Patient Confidentiality When Sharing Medical Data Requires a Symbiotic Relationship Between Technology and Policy*. Artificial Intelligence Laboratory, Massachusetts Institute of Technology, AIWP-WP344, May 1997
L. Sweeney. Weaving Technology and Policy Together to Maintain Confidentiality. *Journal of Law, Medicine & Ethics*, 25, nos. 2&3 (1997): 98-110.

94043
Male
11/29/1976

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


A
B
N
0 0.5 1 2
Kilometers

Computational Disclosure Control

- 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
- How to estimate "external information"?
- **Every** release becomes more external info.

M. Fredrikson, E. Lantz, S. Jha, S. Lin, D. Page, and T. Ristenpart. Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing. In USENIX Security Symposium, pages 17–32, 2014.

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, µ-Argus (Netherlands), k-Similar
- T. Lasko: spectral anonymization
  - Build a model of data that captures the n-th order statistics of the distribution
  - Synthesize “fake” patients from that distribution
- J. Ghosh: detailed modeling of data
  - Build a Bayesian Network model that captures the dependencies among data
  - Synthesize “fake” patients or directly use the model
- Practical approaches:
  - Put data in a secure data enclave for R&D

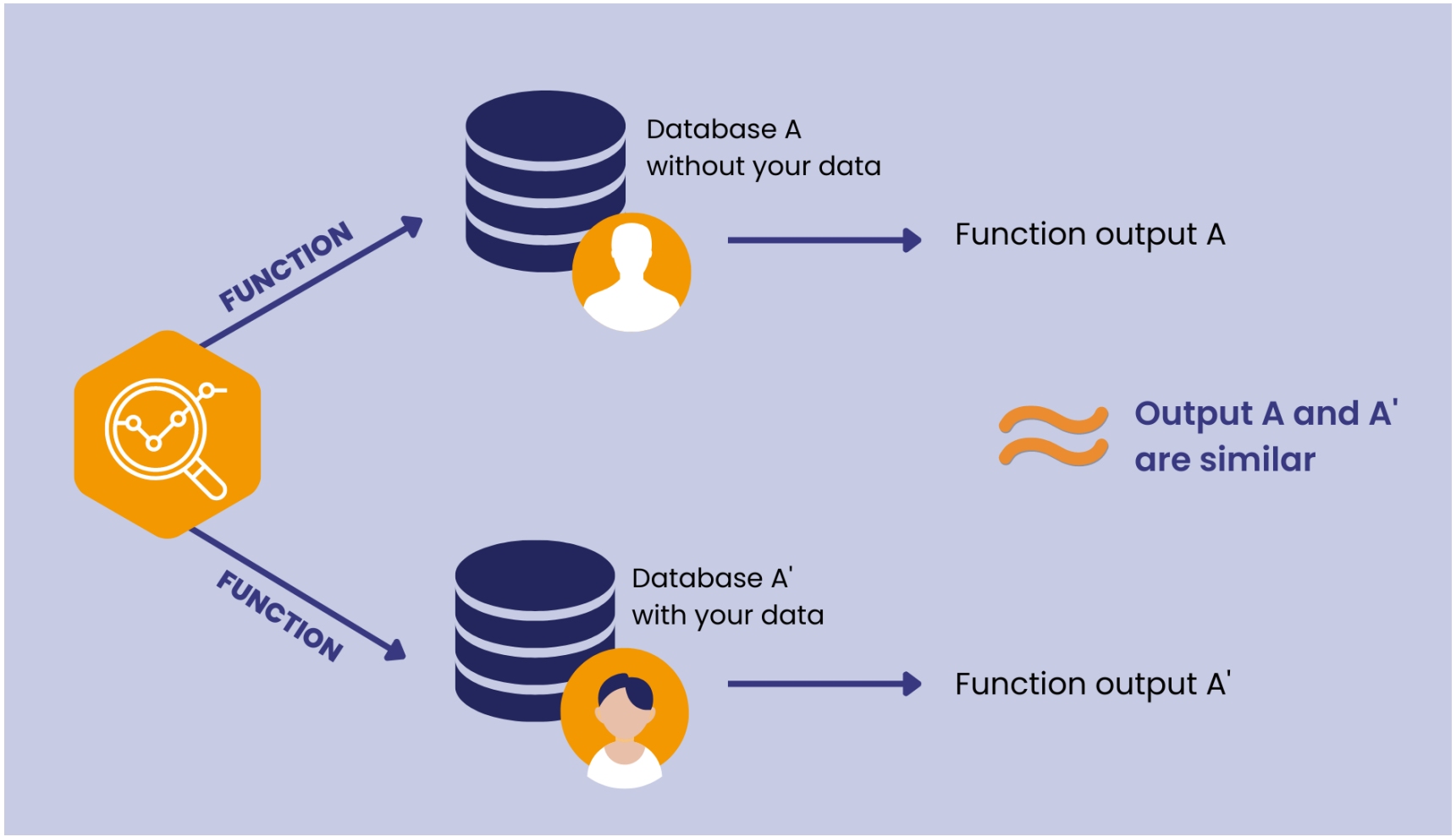
- Differential Privacy

HIPAA complicates patient care

"In this national survey of clinical scientists, only a quarter perceived that the rule has enhanced participants' confidentiality and privacy, whereas the **HIPAA Privacy Rule was perceived to have a substantial, negative influence on the conduct of human subjects health research**, often adding uncertainty, cost, and delay."

https://jamanetwork.com/journals/jama/fullarticle/209447

Differential Privacy



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

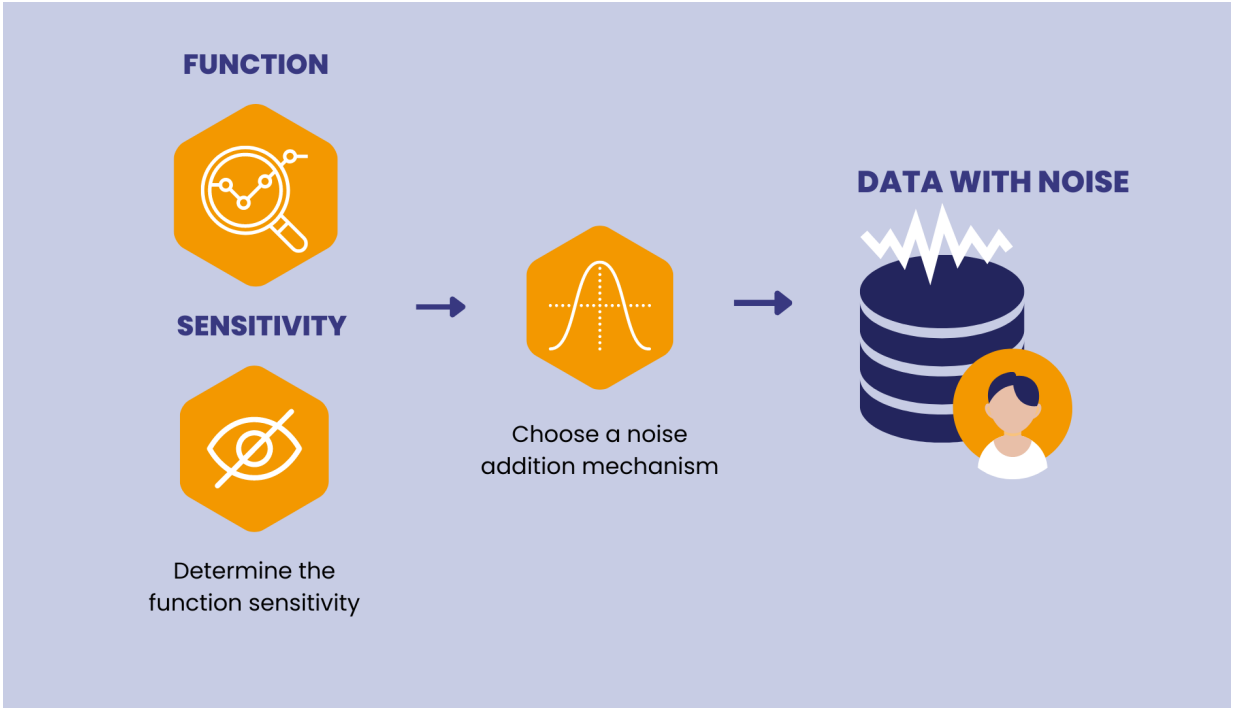
https://www.statice.ai/post/what-is-differential-privacy-definition-mechanisms-examples

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 $S \subset \text{Range}(\mathscr{A})$
  - $\Pr[\mathscr{A}(D_1) \in S] \leq \exp(\epsilon) \cdot \Pr[\mathscr{A}(D_2) \in S] + \delta$
  - $\mathscr{A}$ 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
  - ϵ and δ are (small) numbers; δ is prob. that privacy guarantee fails
- This condition can be achieved for all pairs of D_1, D_2 by having $\mathscr{A}$ 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 ϵ
  - "Privacy Budget"

Dwork C, McSherry F, Nissim K, Smith A. Calibrating Noise to Sensitivity in Private Data Analysis. In Springer Berlin Heidelberg; 2006. p. 265–84.

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 $\{x_1, \ldots, x_N\}$, loss function $\mathcal{L}(\theta) = \frac{1}{N}\sum_i \mathcal{L}(\theta, x_i)$. Parameters: learning rate η_t, noise scale σ, group size L, gradient norm bound C.

Initialize θ_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

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

Add noise

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

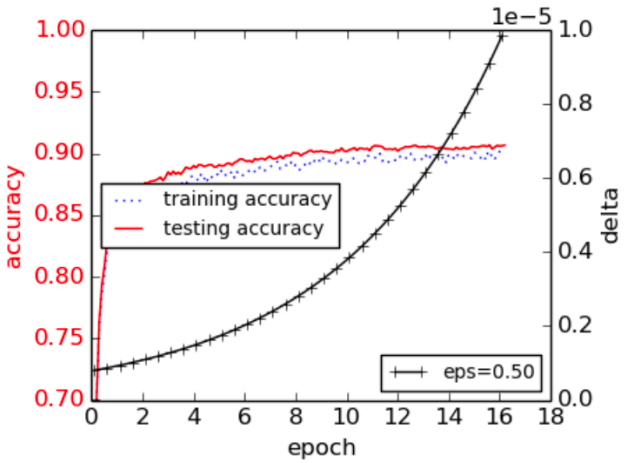
Descent

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

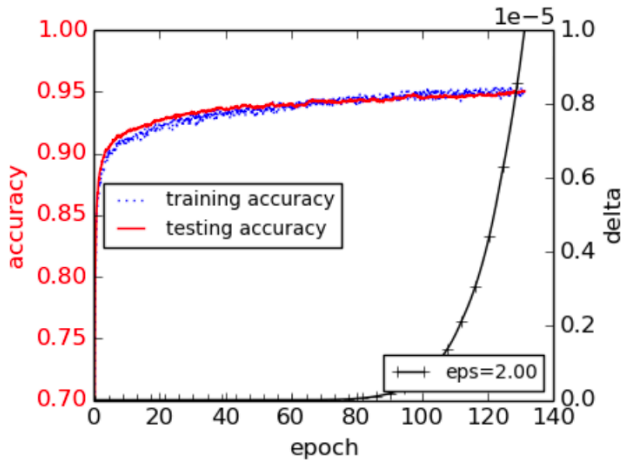
Output θ_T and compute the overall privacy cost (ε, δ) using a privacy accounting method.

Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, et al. Deep Learning with Differential Privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security [Internet]. 2016 [cited 2023 Apr 11]. p. 308–18. Available from: http://arxiv.org/abs/1607.00133

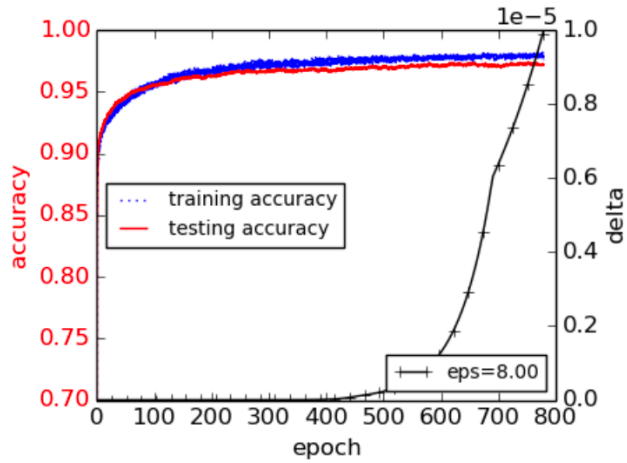
Works Well for 1-layer Model for MNIST (less well on CIFAR-10)



(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 (σ, σ_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.

Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, Talwar K, et al. Deep Learning with Differential Privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security [Internet]. 2016 [cited 2023 Apr 11]. p. 308–18. Available from: http://arxiv.org/abs/1607.00133

The privacy-utility trade-off

- Evaluate using the following datasets:

Dataset	Data Type	Outcome Variable	n	d	Classification Task	Tail Size	Protected Attributes	Evaluation
HEALTH CARE								
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
Vision Baselines								
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.

Suriyakumar, Papernot, Goldenberg, Ghassemi. Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings, FAccT '21

The privacy-utility trade-off

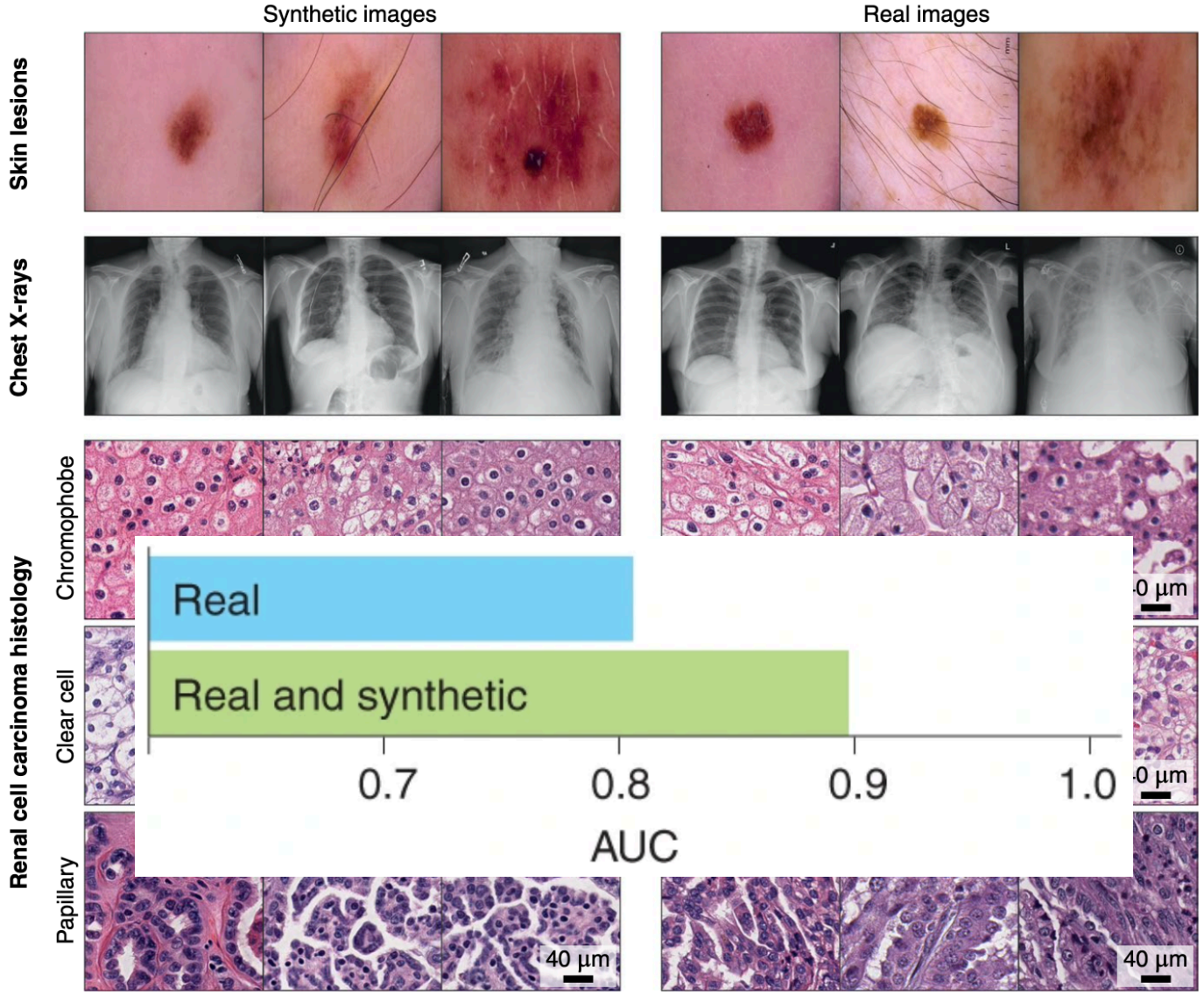
Vision Baselines				
Dataset	Model	None (ϵ, δ)	Low (ϵ, δ)	High (ϵ, δ)
MNIST	CNN	98.83 ± 0.06 $(\infty, 0)$	98.58 ± 0.06 $(2.6 \cdot 10^5)$	93.78 ± 0.25 (2.01)
FashionMNIST	CNN	87.92 ± 0.19 $(\infty, 0)$	87.90 ± 0.16 $(2.6 \cdot 10^5)$	79.53 ± 0.10 (2.01)
MIMIC-III				
Task	Model	None (ϵ, δ)	Low (ϵ, δ)	High (ϵ, δ)
Mortality	LR	0.82 ± 0.03 $(\infty, 0)$	0.76 ± 0.05 $(3.50 \cdot 10^5, 10^{-5})$	0.60 ± 0.04 $(3.54, 10^{-5})$
	GRUD	0.79 ± 0.03 $(\infty, 0)$	0.59 ± 0.09 $(1.59 \cdot 10^5, 10^{-5})$	0.53 ± 0.03 $(2.65, 10^{-5})$
Length of Stay > 3	LR	0.69 ± 0.02 $(\infty, 0)$	0.66 ± 0.03 $(3.50 \cdot 10^5, 10^{-5})$	0.60 ± 0.04 $(3.54, 10^{-5})$
	GRUD	0.67 ± 0.03 $(\infty, 0)$	0.63 ± 0.02 $(1.59 \cdot 10^5, 10^{-5})$	0.61 ± 0.03 $(2.65, 10^{-5})$
Intervention Onset (Vaso)	LR	0.90 ± 0.03 $(\infty, 0)$	0.87 ± 0.03 $(1.63 \cdot 10^7, 10^{-5})$	0.77 ± 0.05 $(0.94, 10^{-5})$
	CNN	0.88 ± 0.04 $(\infty, 0)$	0.86 ± 0.02 $(5.95 \cdot 10^7, 10^{-5})$	0.68 ± 0.04 $(0.66, 10^{-5})$
NIH Chest X-Ray				
Metric	Model	None (ϵ, δ)	Low (ϵ, δ)	High (ϵ, δ)
Average AUC	DenseNet-121	0.84 ± 0.00 $(\infty, 0)$	0.51 ± 0.01 $(1.74 \cdot 10^5, 10^{-6})$	0.49 ± 0.00 $(0.84, 10^{-6})$
Best AUC	DenseNet-121	0.98 ± 0.00 (Hernia)	0.54 ± 0.04 (Edema)	0.54 ± 0.05 (Pleural Thickening)
Worst AUC	DenseNet-121	0.72 ± 0.00 (Infiltration)	0.48 ± 0.02 (Fibrosis)	0.47 ± 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 ϵ, δ guarantees in parentheses, where ϵ represents the privacy loss (lower is better) and δ represents the probability that the guarantee does not hold (lower is better).

Thought on Differential Privacy in Healthcare

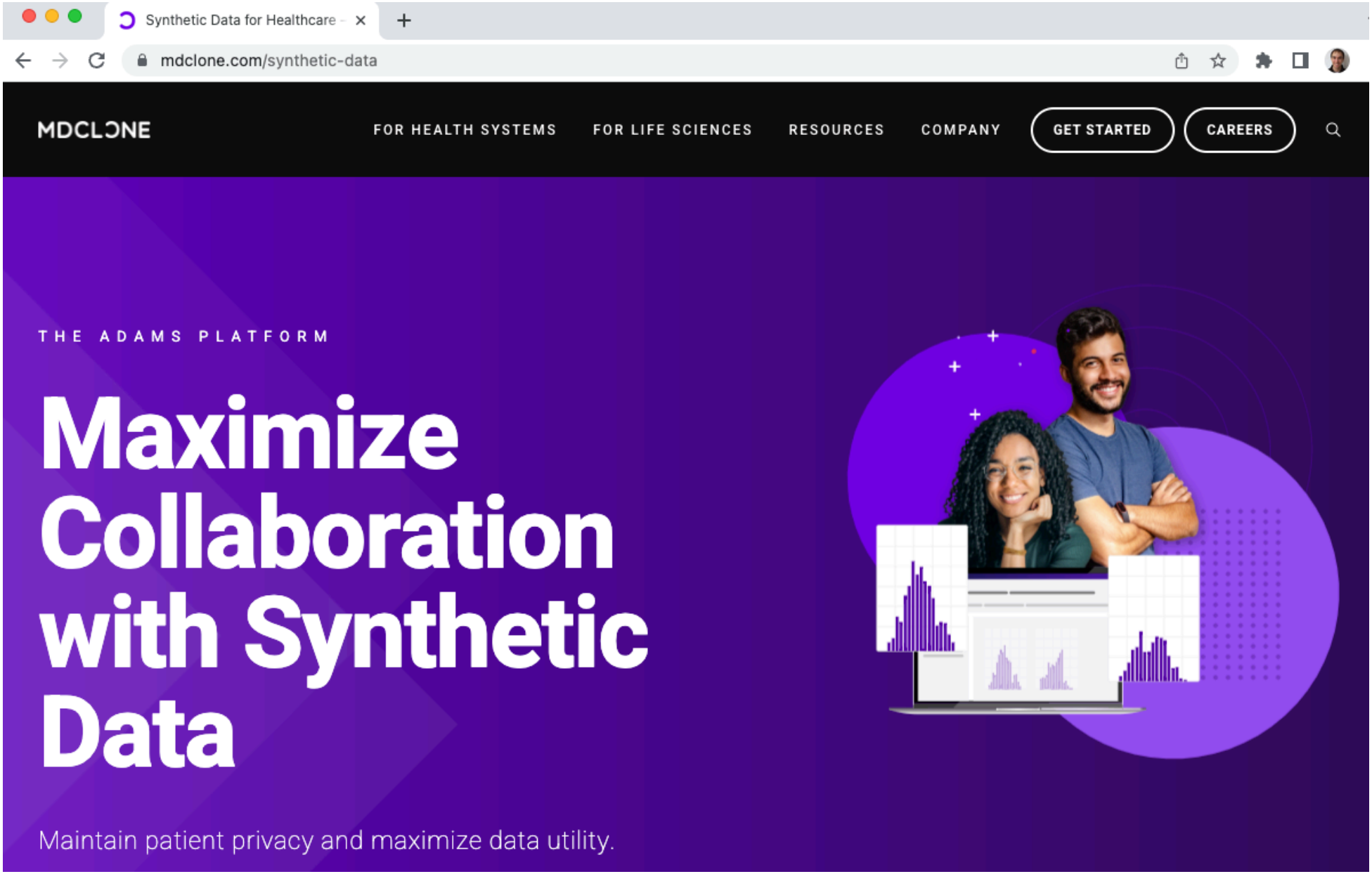
- Great to quantify privacy risks
  - though setting and interpretation of ϵ remains problematic
- Very high cost in model performance
- DP can serve as a regularizer, thought to improve performance with dataset shift
  - Evidence from biomedical fields does not support this hope
- By washing out tails,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

Synthetic Data Generation by GANs



Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng [Internet]. 2021 Jun 15;5(6):493–7. Available from: https://www.nature.com/articles/s41551-021-00751-8

Synthetic data generation



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.

function MI-FACE($label, \alpha, \beta, \gamma, \lambda$)
  $c(\mathbf{x}) \stackrel{\text{def}}{=} 1 - \tilde{f}_{label}(\mathbf{x})$
  $\mathbf{x}_0 \leftarrow \mathbf{0}$
  for $i \leftarrow 1 \ldots \alpha$ **do**
    $\mathbf{x}_i \leftarrow \text{PROCESS}(\mathbf{x}_{i-1} - \lambda \cdot \nabla c(\mathbf{x}_{i-1}))$
    if $c(\mathbf{x}_i) \geq \max(c(\mathbf{x}_{i-1}), \ldots, c(\mathbf{x}_{i-\beta}))$ **then**
      break
    if $c(\mathbf{x}_i) \leq \gamma$ **then**
      break
  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).

M. Fredrikson, S. Jha, T. Ristenpart. Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures. CCS '15.

Federated Learning in Large Clinical Research Networks

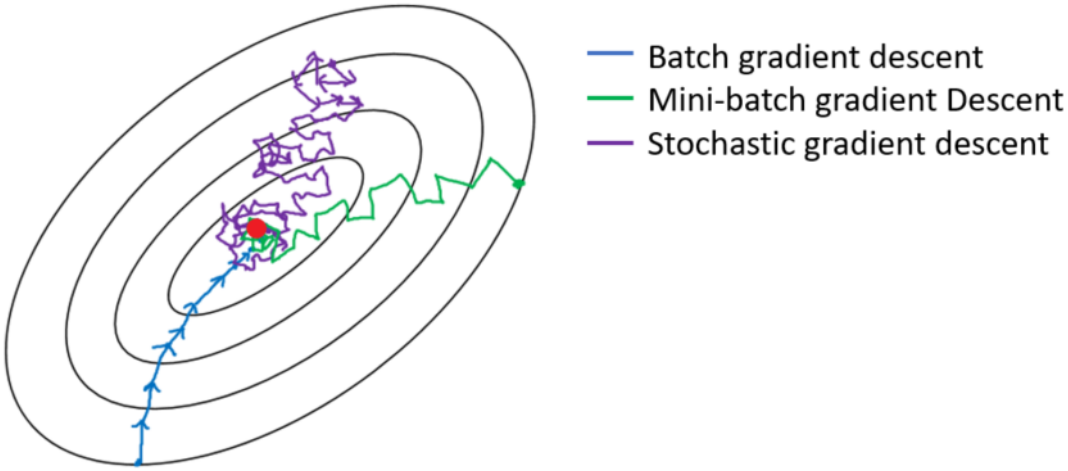
Fei Wang
Associate Professor
Department of Population Health Sciences
Weill Cornell Medicine
feiwang.cornell@gmail.com
@feiwang03
https://wcm-wanglab.github.io/index.html

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

Stochastic Gradient Descent

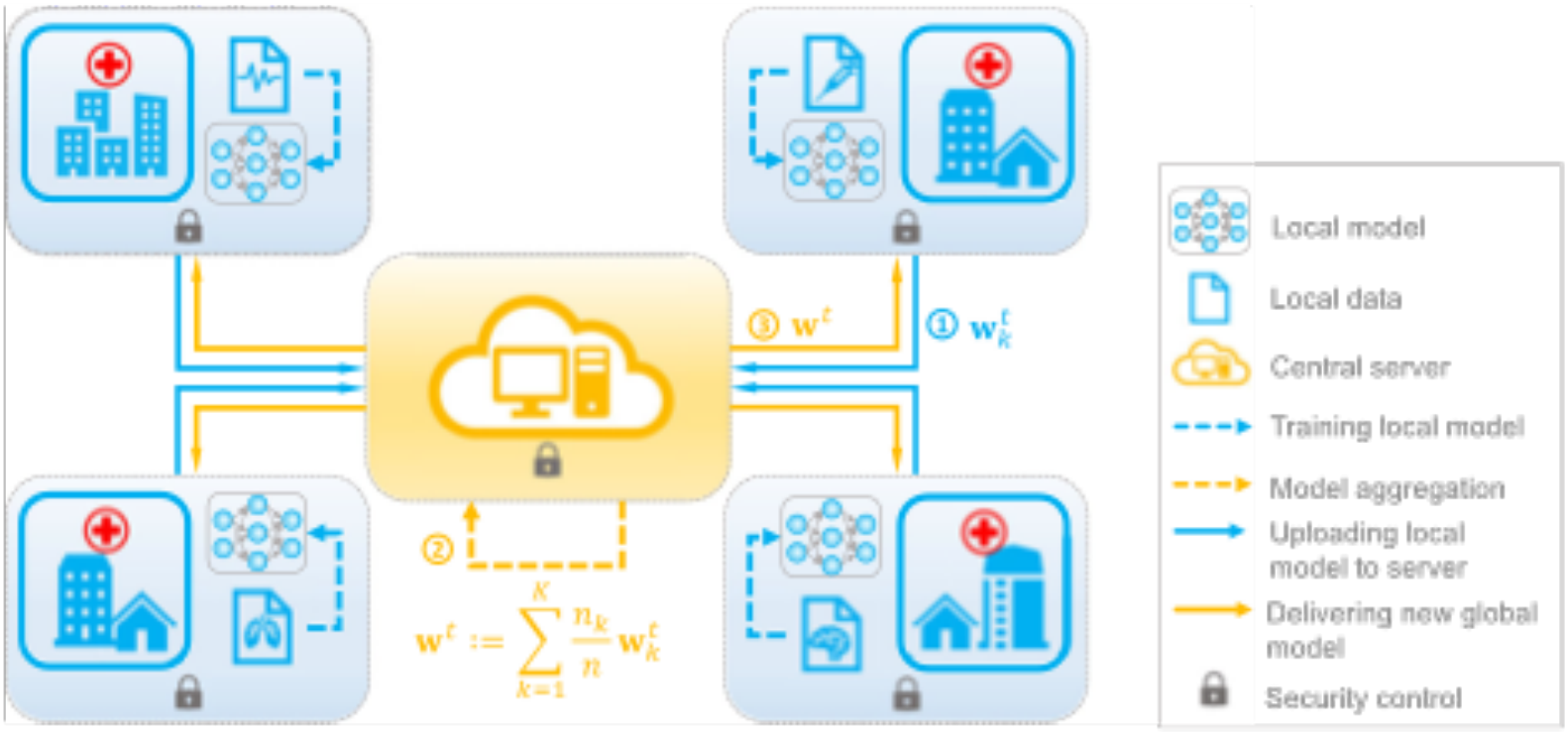
- At each step of gradient descent, instead of compute for all training samples, randomly pick a small subset (mini-batch) of training samples

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



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

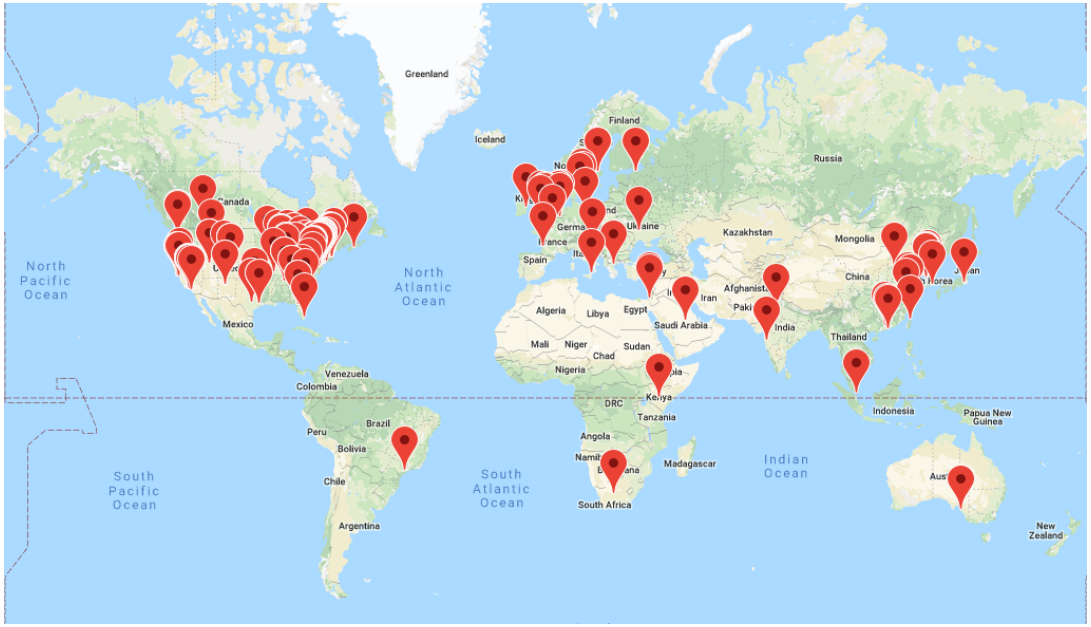
Federated Learning

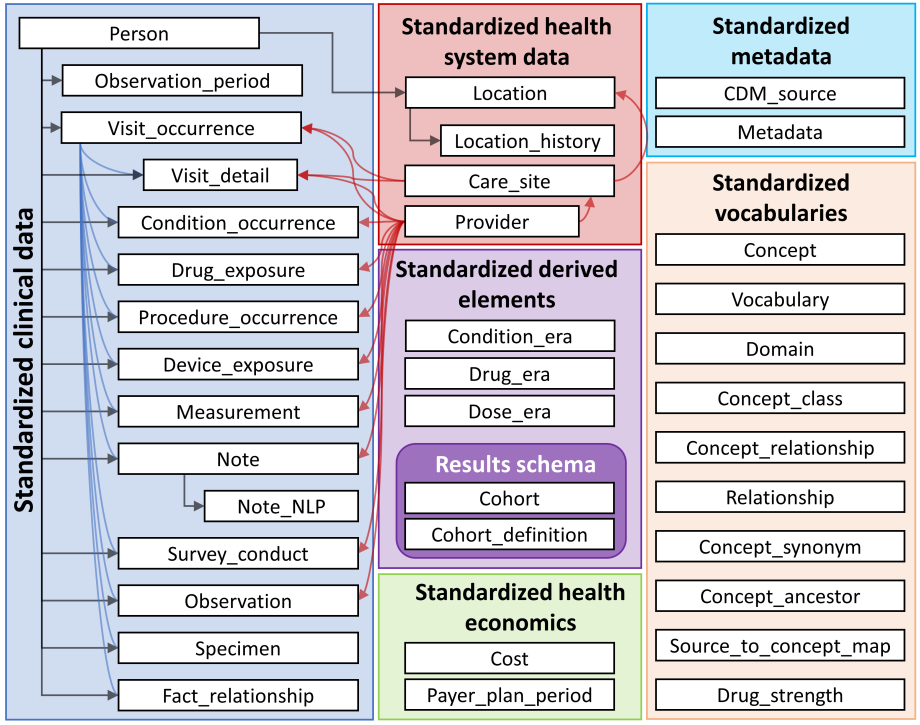


Xu, Jie, Benjamin S. Glicksberg, Chang Su, Peter Walker, Jiang Bian, and ***Fei Wang***. "Federated learning for healthcare informatics." *Journal of Healthcare Informatics Research* (2020): 1-19.

Clinical Research Networks

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS





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

https://inst.eecs.berkeley.edu/~cs294-163/fa19/slides/federated-learning.pdf

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 η 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

McMahan, Brendan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. "Communication-efficient learning of deep networks from decentralized data." In *Artificial intelligence and statistics*, pp. 1273-1282. PMLR, 2017.

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	**0.766** (0.763-0.768)	**0.810** (0.808-0.812)

*Area under the receiver operating characteristic curve

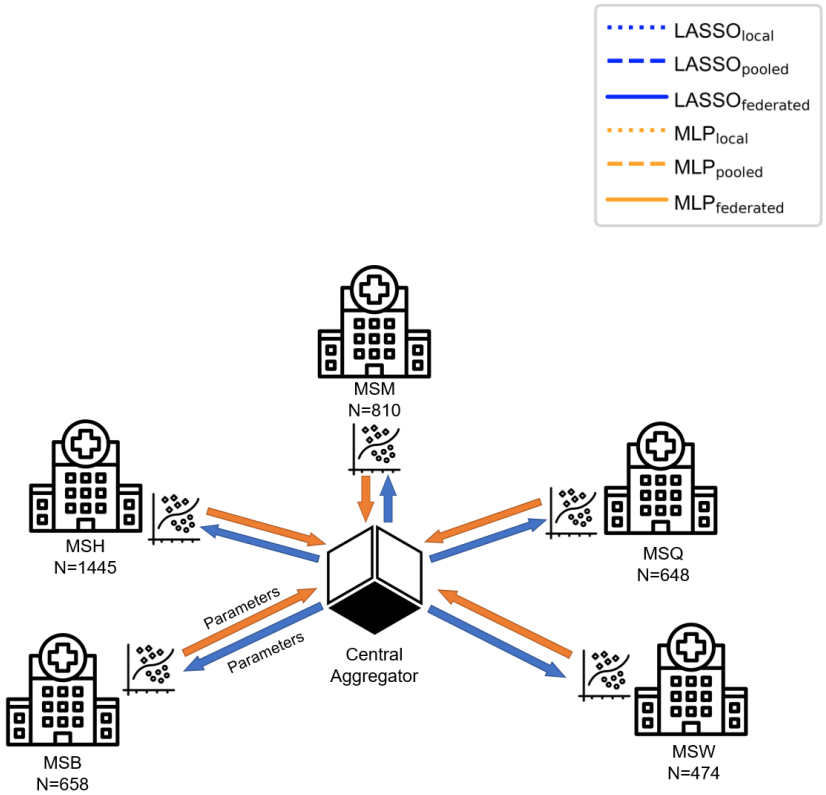
Summary: Federated model classifiers outperform locally trained classifiers in predicting mortality among hospitalized patients with COVID-19.

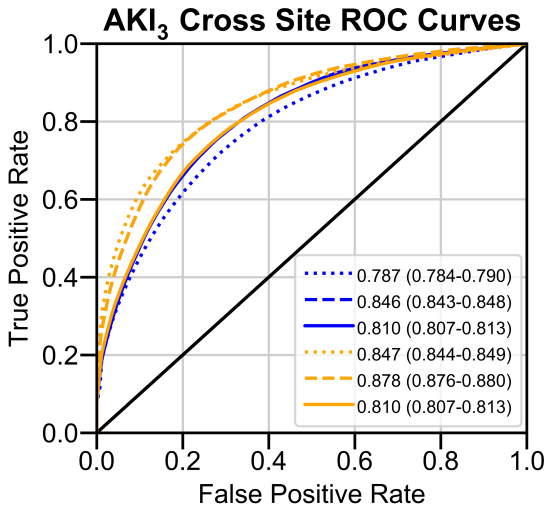
Akhil Vaid, Suraj K Jaladanki, Jie Xu, Shelly Teng, Arvind Kumar, Samuel Lee, Sulaiman Somani, Ishan Paranjpe, Jessica K De Freitas, Tingyi Wanyan, Kipp W Johnson, Mesude Bicak, Eyal Klang, Young Joon Kwon, Anthony Costa, Shan Zhao, Riccardo Miotto, Alexander W Charney, Erwin Böttinger, Zahi A Fayad, 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.

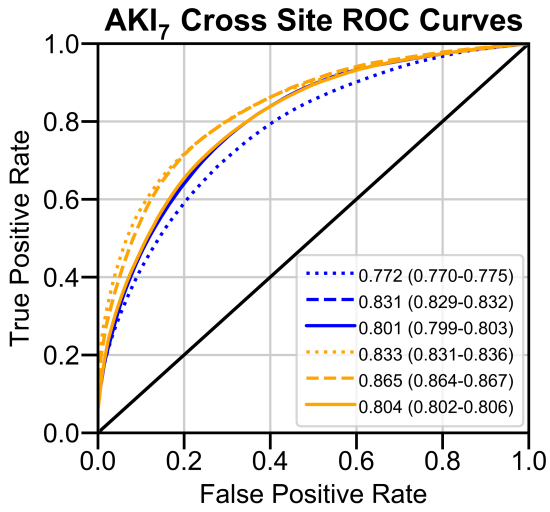
Characteristic	Mount Sinai Brooklyn	Mount Sinai Hospital	Mount Sinai Morningside	Mount Sinai Queens	Mount Sinai West	*P* value
Number of patients, n	611	1644	749	540	485	_[b]
Gender, n (%)						
Male	338 (55.3)	951 (57.8)	411 (54.9)	344 (63.7)	257 (53.0)	.004
Female	273 (44.7)	693 (42.2)	338 (45.1)	196 (36.3)	228 (47.0)	.004
Age (years), median (IQR)	72.5 (63.6-82.7)	63.3 (51.3-73.2)	69.8 (57.4-80.3)	68.1 (57.1-78.8)	66.3 (52.5-77.6)	<.001
Ethnicity, n (%)						
Hispanic	21 (3.4)	460 (28.0)	259 (34.6)	198 (36.7)	111 (22.9)	<.001
Non-Hispanic	416 (68.1)	892 (54.3)	452 (60.3)	287 (53.1)	349 (72.0)	<.001
Unknown	174 (28.5)	292 (17.8)	38 (5.1)	55 (10.2)	25 (5.2)	<.001
Race, n (%)						
Asian	13 (2.1)	83 (5.0)	16 (2.1)	56 (10.4)	27 (5.6)	<.001
Black/African American	323 (52.9)	388 (23.6)	266 (35.5)	64 (11.9)	109 (22.5)	<.001
Other	54 (8.8)	705 (42.9)	343 (45.8)	288 (53.3)	164 (33.8)	<.001
Unknown	27 (4.4)	87 (5.3)	25 (3.3)	14 (2.6)	14 (2.9)	<.001
White	194 (31.8)	381 (23.2)	99 (13.2)	118 (21.9)	171 (35.3)	<.001

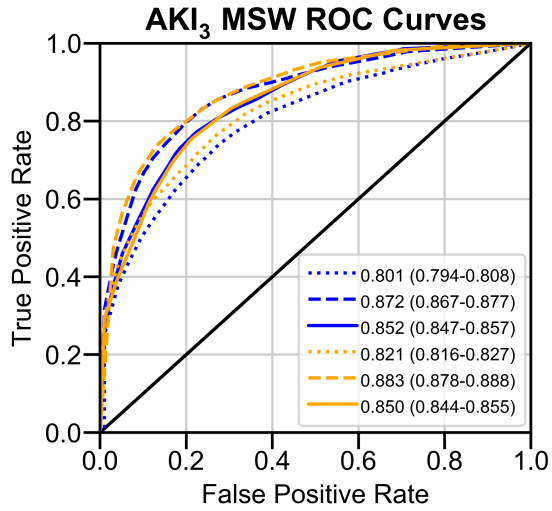
Model	Mount Sinai Brooklyn (n=611), AUROC (95% CI)	Mount Sinai Hospital (n=1644), AUROC (95% CI)	Mount Sinai Morningside (n=749), AUROC (95% CI)	Mount Sinai Queens (n=540), AUROC (95% CI)	Mount Sinai West (n=485), AUROC (95% CI)
LASSO model					
Local	0.791 (0.788-0.795)	0.693 (0.689-0.696)	0.66 (0.656-0.664)	0.706 (0.702-0.710)	0.482 (0.473-0.491)
Pooled	0.816 (0.814-0.819)	0.791 (0.788-0.794)	0.789 (0.785-0.792)	0.734 (0.730-0.737)	0.829 (0.824-0.834)
Federated	0.793 (0.790-0.796)	0.772 (0.769-0.774)	0.767 (0.764-0.771)	0.694 (0.690-0.698)	0.801 (0.796-0.807)
MLP model					
Local	0.822 (0.820-0.825)	0.750 (0.747-0.754)	0.747 (0.743-0.751)	0.791 (0.788 - 0.795)	0.719 (0.711-0.727)
Pooled	0.823 (0.820-0.826)	0.792 (0.789-0.795)	0.751 (0.747-0.755)	0.783 (0.779-0.786)	0.842 (0.837-0.847)
Federated (no noise	0.829 (0.826-0.832)	0.786 (0.782-0.789)	0.791 (0.788-0.795)	0.809 (0.806-0.812)	0.836 (0.83-0.841)

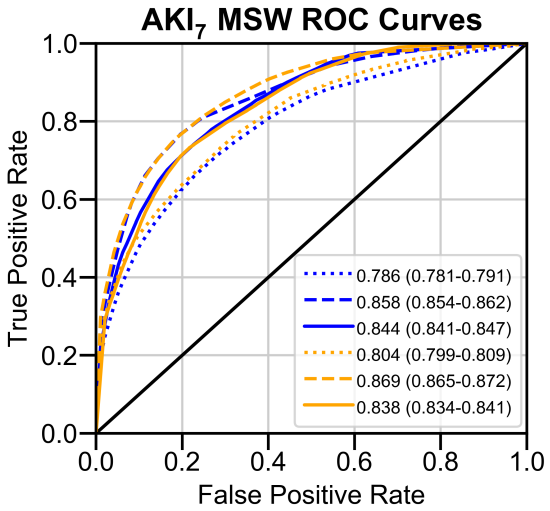
Acute Kidney Injury in COVID-19









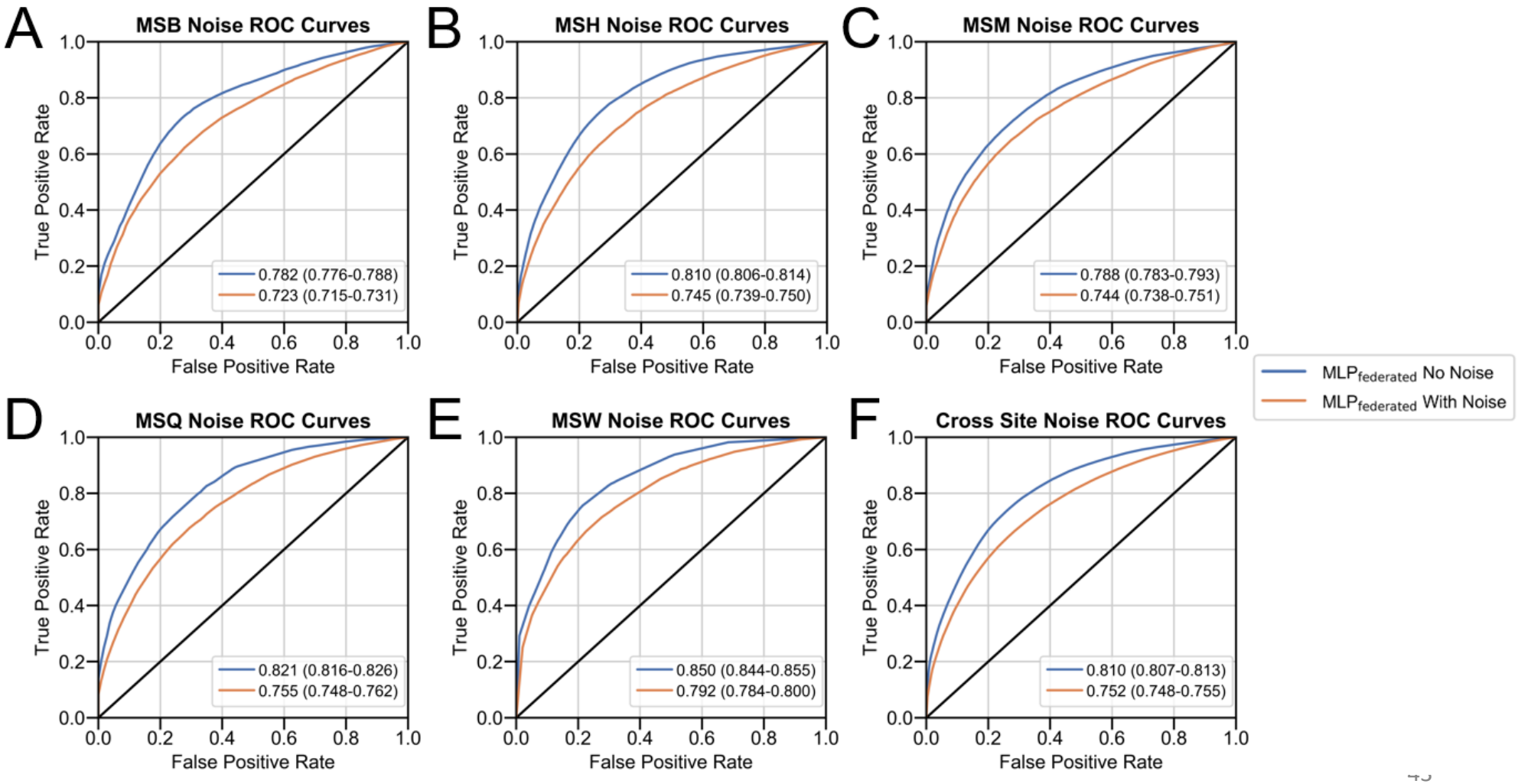


Jaladanki, Suraj K., Akhil Vaid, Ashwin S. Sawant, Jie Xu, Kush Shah, Sergio Dellepiane, Ishan Paranjpe, Lili Chan, Alexander W Charney, **Fei Wang**, Benjamin S Glicksberg, Karandeep Singh, Girish N Nadkarn "Development of a federated learning approach to predict acute kidney injury in adult hospitalized patients with COVID-19 in New York City." *medRxiv* (2021).

Prediction performance of federated models was generally higher than single-hospital models and was comparable to pooled-data models.

Most Important Features at MSW for AKI_3

- $LASSO_{local}$
  - history of stroke
  - Black race
  - Hispanic/Latino ethnicity
- $LASSO_{federated}$
  - Blood urea nitrogen
  - age
  - albumin

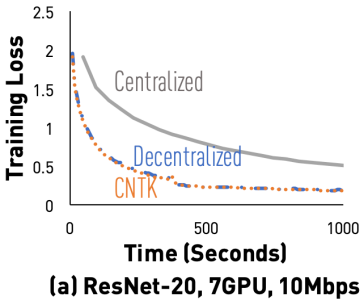

A
MSB Noise ROC Curves
True Positive Rate
False Positive Rate
0.782 (0.776-0.788)
0.723 (0.715-0.731)
B
MSH Noise ROC Curves
0.810 (0.806-0.814)
0.745 (0.739-0.750)
C
MSM Noise ROC Curves
0.788 (0.783-0.793)
0.744 (0.738-0.751)
D
MSQ Noise ROC Curves
0.821 (0.816-0.826)
0.755 (0.748-0.762)
E
MSW Noise ROC Curves
0.850 (0.844-0.855)
0.792 (0.784-0.800)
F
Cross Site Noise ROC Curves
0.810 (0.807-0.813)
0.752 (0.748-0.755)
MLP federated No Noise
MLP federated With Noise

Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent

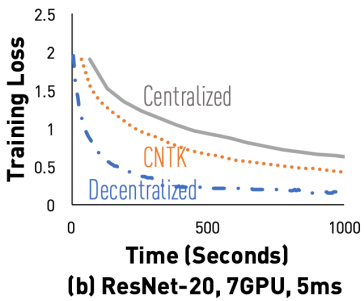
- Computational complexity of C-PSGD same as for D-PSGD
- But, lower communication cost on the busiest node

Algorithm	communication complexity on the busiest node	computational complexity
C-PSGD (mini-batch SGD)	$O(n)$	$O(\frac{n}{\epsilon} + \frac{1}{\epsilon^2})$
D-PSGD	$O(\text{Deg(network)})$	$O(\frac{n}{\epsilon} + \frac{1}{\epsilon^2})$

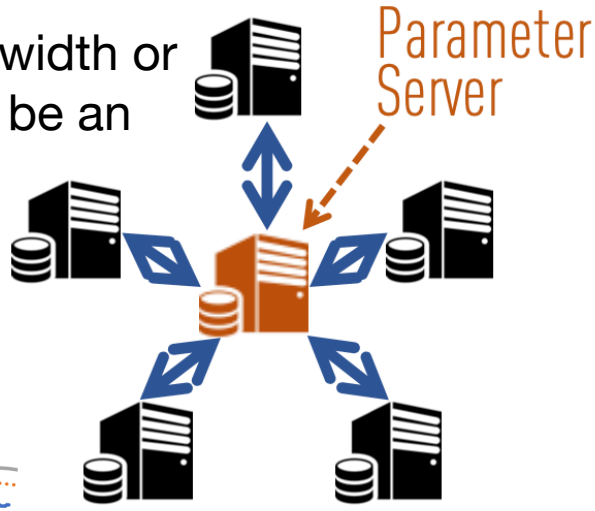
- Experiments on up to 112 GPUs
- In networks with low bandwidth or high latency, D-PSGD can be an order of magnitude faster



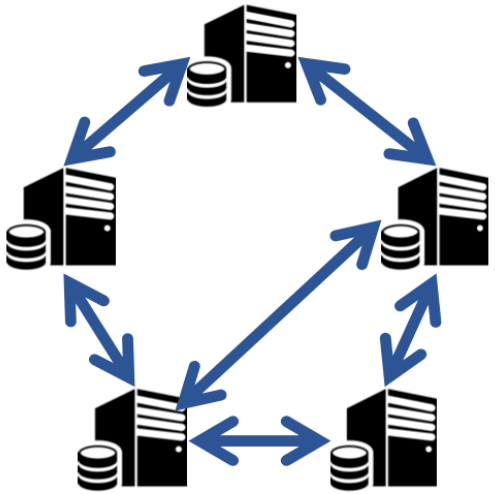
(a) ResNet-20, 7GPU, 10Mbps



(b) ResNet-20, 7GPU, 5ms



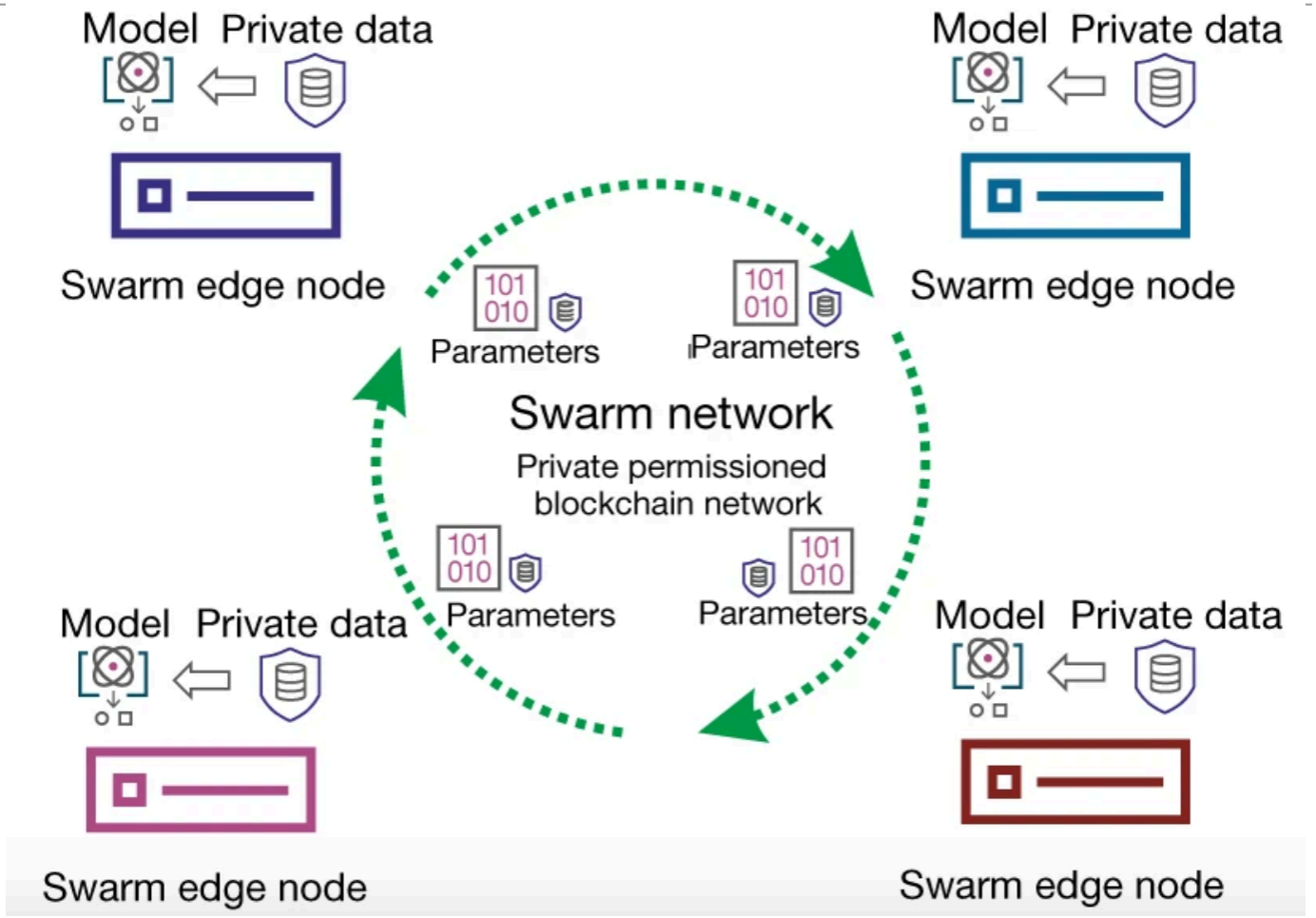
Centralized Topology


(b) Decentralized Topology

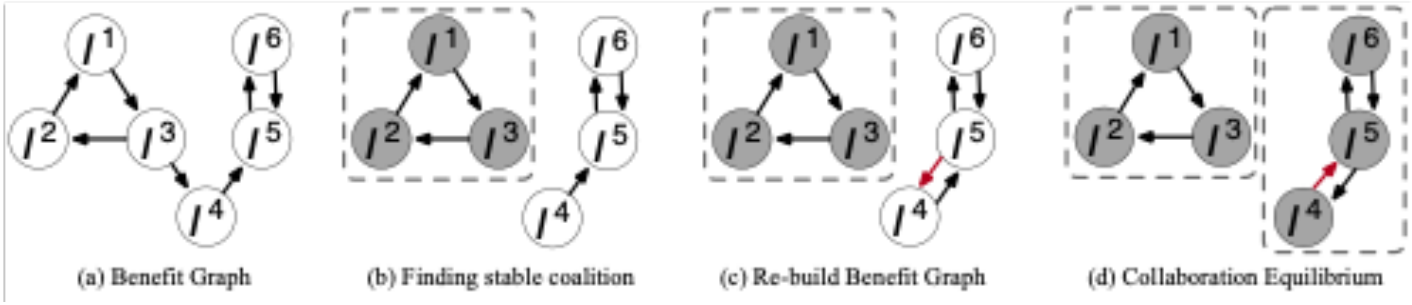
Lian X, Zhang C, Zhang H, Hsieh CJ, Zhang W, Liu J. Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent [Internet]. arXiv; 2017. Available from: http://arxiv.org/abs/1705.09056

Swarm Learning

Warnat-Herresthal, Stefanie, Hartmut Schultze, Krishnaprasad Lingadahalli Shastry, Sathyanarayanan Manamohan, Saikat Mukherjee, Vishesh Garg, Ravi Sarveswara et al. "Swarm Learning for decentralized and confidential clinical machine learning." *Nature* 594, no. 7862 (2021): 265-270.

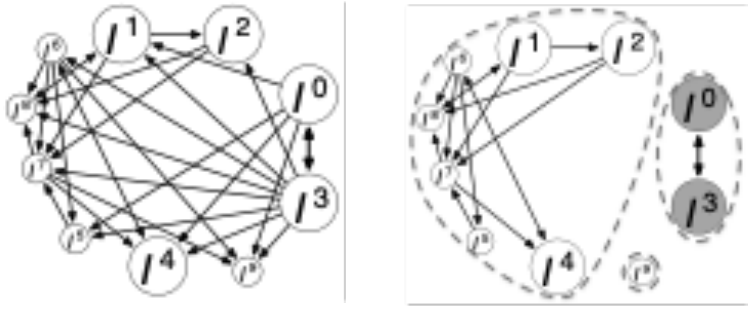


Learning to Collaborate



Algorithm 1: Achieving collaboration equilibrium

Input: N institutions $\boldsymbol{I} = \{I^i\}_{i=1}^N$ seeking collaborating with others
Set original client set $C \leftarrow \boldsymbol{I}$;
Set collaboration strategy $S \leftarrow \emptyset$;
while $C \neq \emptyset$ **do**
  forall *client* $I^i \in C$ **do**
    Determine the OCS of I^i by SPO;
  Construct the benefit graph $BG(C)$;
  Search for all strongly connected components $\{C^1, C^2, ...C^k\}$ of $BG(C)$ using **Tarjan** algorithm;
  forall i = *1, 2, 3,... k* **do**
    if C^i *is stable coalition* **then**
      $C \leftarrow C \backslash C^i$;
      $S \leftarrow S \cup \{C^i\}$;
Output: collaboration strategy S



Sen Cui, Jian Liang, Weishen Pan, Kun Chen, Changshui Zhang, **Fei Wang**. Learning to Collaborate. https://arxiv.org/abs/2108.07926 . 2021.

Federated Learning Conclusions

Clinical problems are typically complicated with limited sample size. Clinical data are sensitive. All these make federated learning important.

- Data standardization/harmonization is important before federated learning can be applied.
- To further protect privacy, differential privacy/block chain techniques could be helpful.
- Incentives/benefits are important to consider for participating in federated learning.
- In addition to model accuracy, model fairness could be important as well.

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

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?