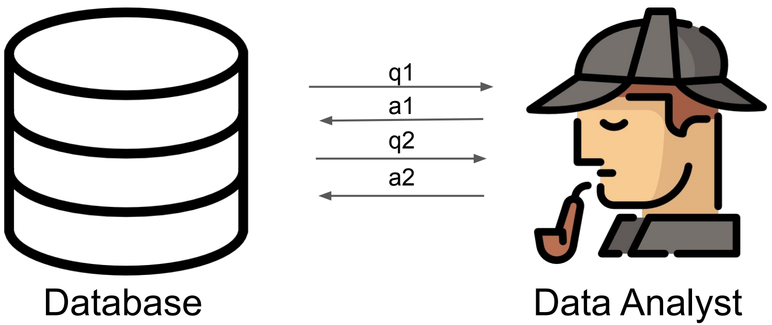
Differential Privacy

by In Woo Park

Analyzing Data While Preserving Privacy



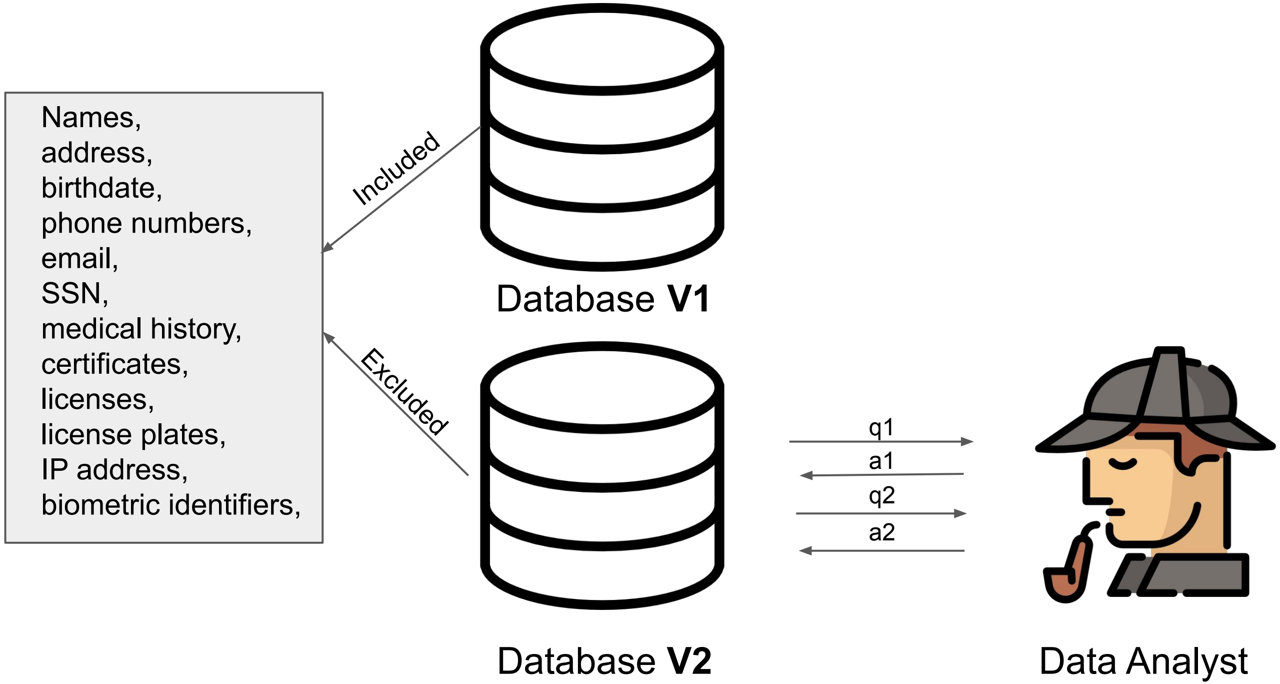
- Census Data
- Epidemic detection based on OTC drug purchases
- Cancer detection based on insurance premiums and smoking

How can we analyze data

while keeping it private?

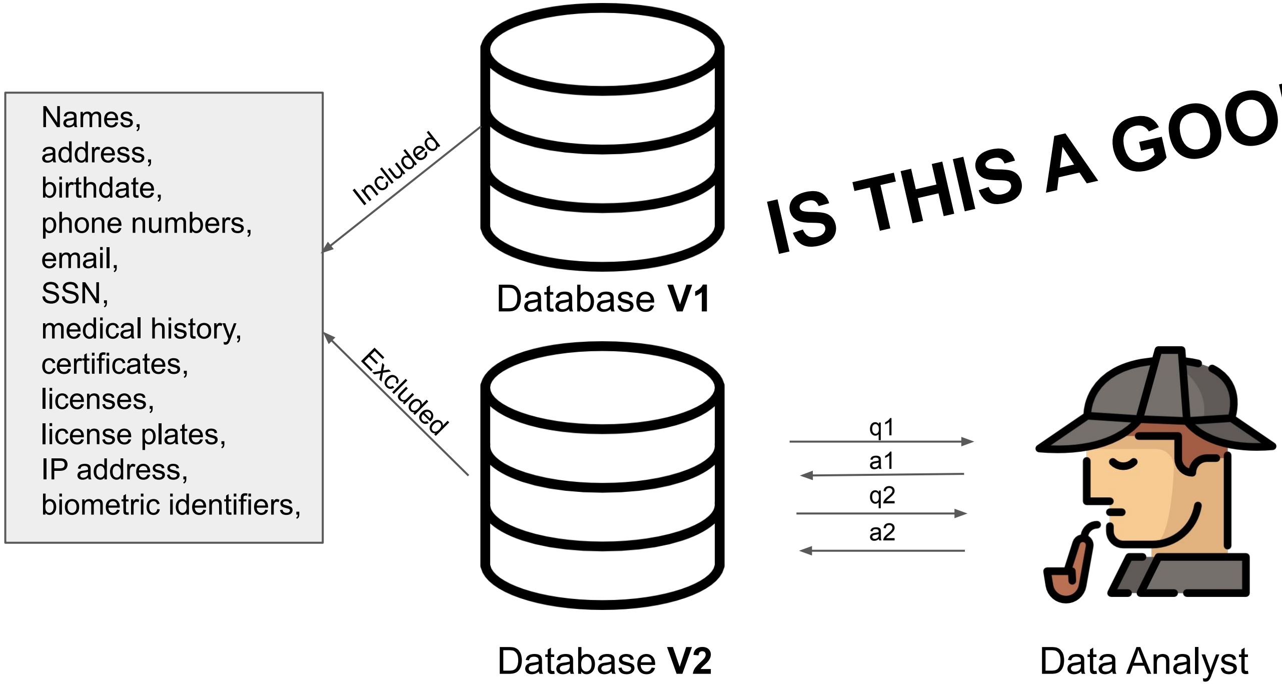
Idea: De-identified Data

- If a data set contains any amount or kind of personal information, it cannot be considered de-identified.



Idea: De-identified Data

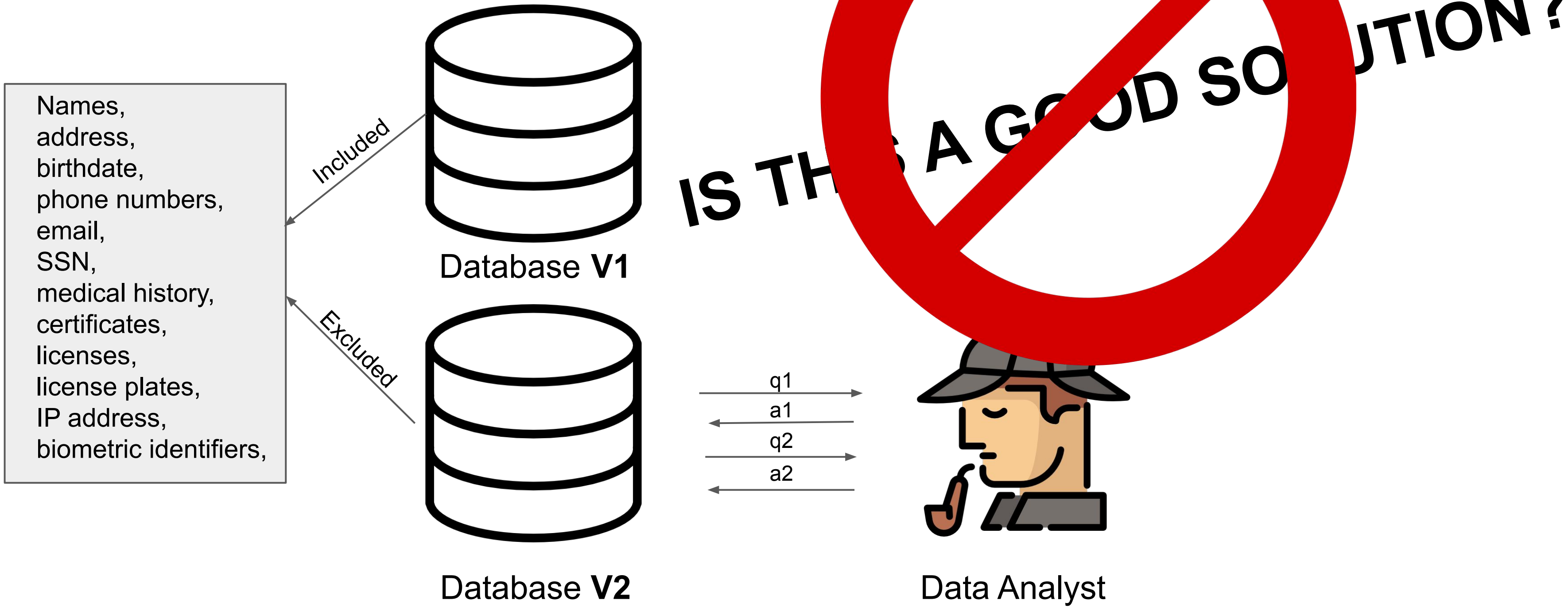
- If a data set contains any amount or kind of personal information, it cannot be considered de-identified.



IS THIS A GOOD SOLUTION?

Idea: De-identified Data

- If a data set contains any amount or kind of persona mation, it cann considered de-identified.

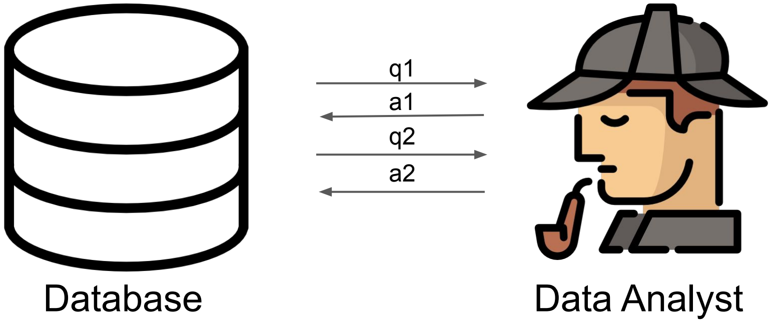


Idea: De-identified Data = 🤔

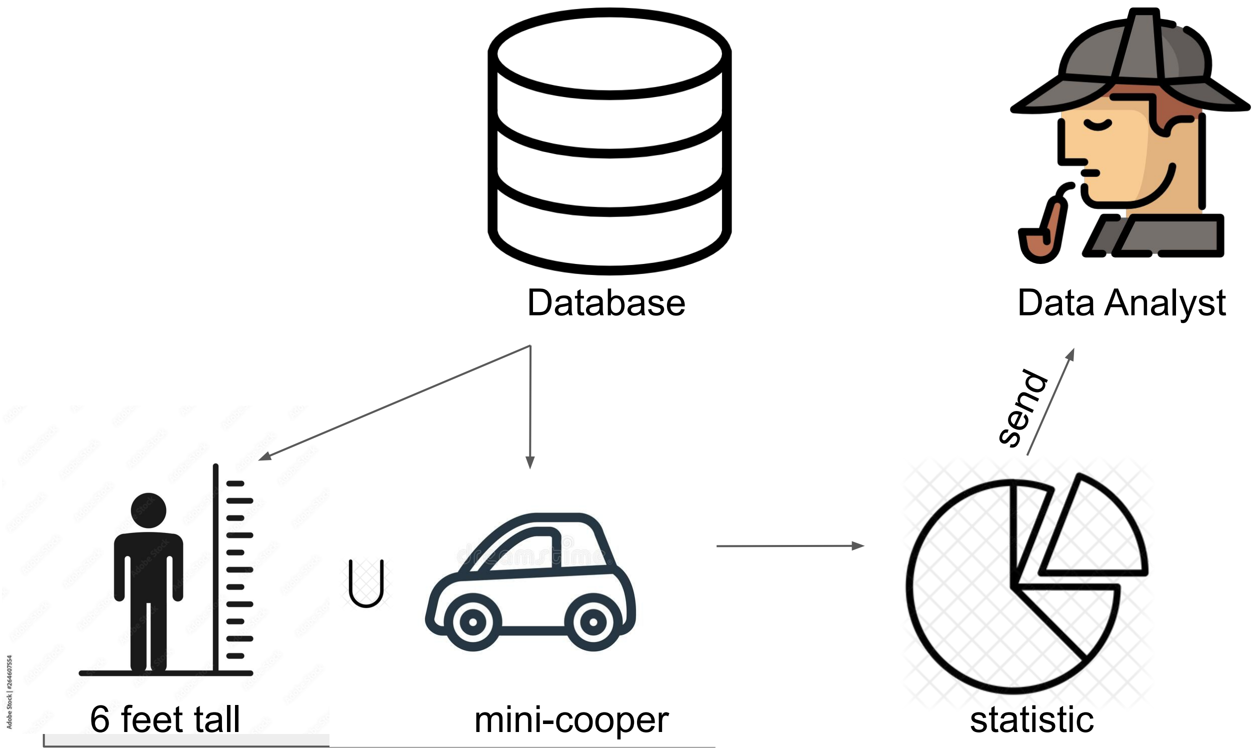
- De-identified != anonymized
  - identifiers are removed, but rest of the data is untouched
  - can still be identified because of other datasets in the world

- (identified dataset) ∩ (de-identified data) = re-identified data
  - NOT GOOD !

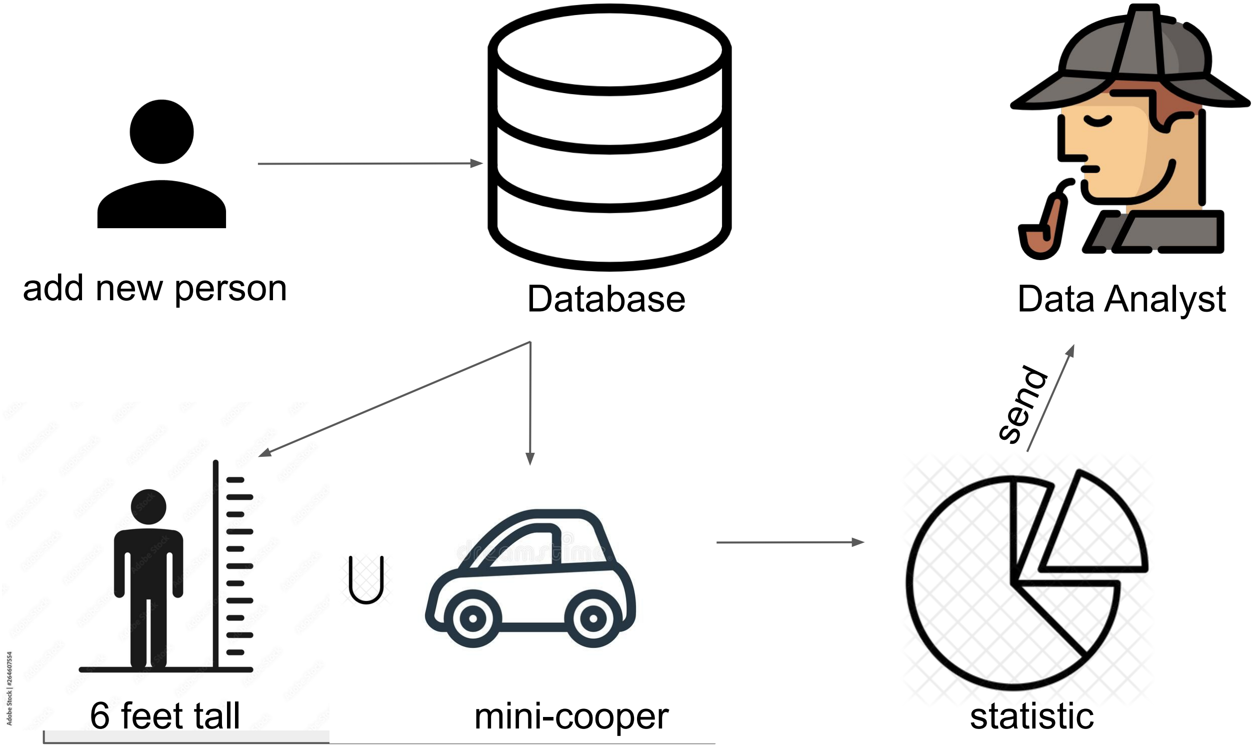
Idea: Just Give Statistics



Idea: Just Give Statistics



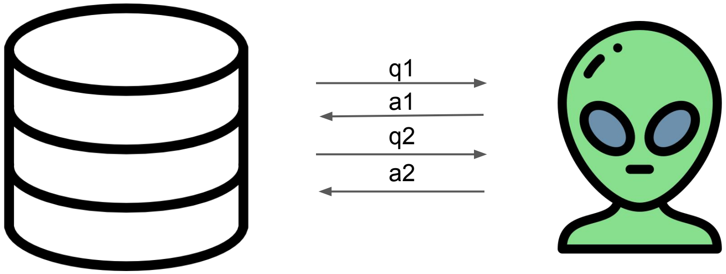
Idea: Just Give Statistics



Idea: Just Give Statistics = 🤔

- Fundamental Law of Info Reconstruction
    - “overly accurate estimates of too many statistics can completely destroy privacy”
    - Dinur and Nissim, 2003; Dwork et al, 2007; Homer et al, 2008, Dwork et al., 2015b
- Findings:
    - “randomness”
    - attacks work so long as the amount of **noise** is small enough
    - attacks fail if the amount of noise is large enough
    - attacks fail even if the amount of noise is small **if** queries are correspondingly small
- Conclusions:
    - As long as limits are placed on queries (relative to amount of noise), the attack fails
    - **This paper pioneered differential privacy!**

Example: Learning From Data

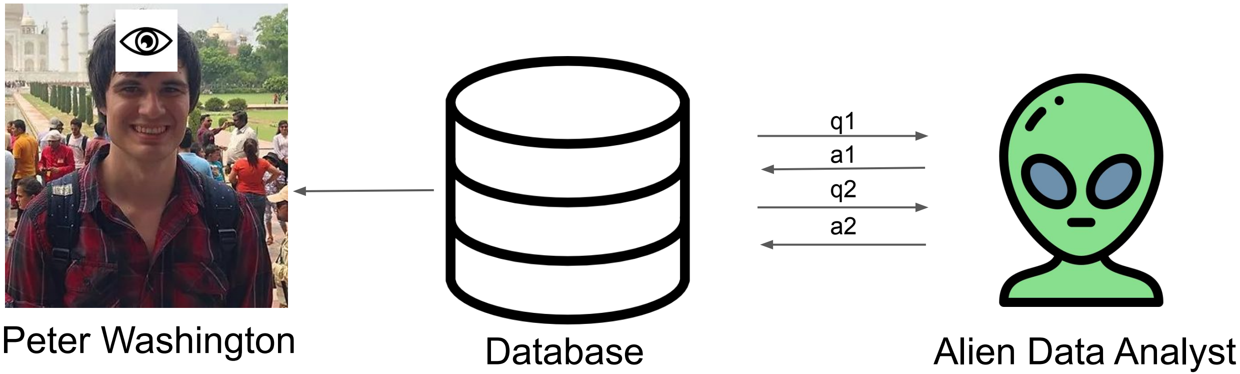


Database

Alien Data Analyst

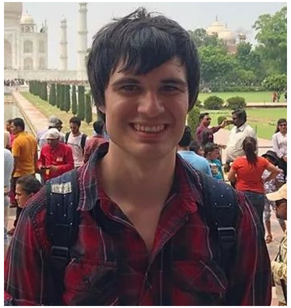
“I think all humans have 3 eyes”

Example: Learning From Data

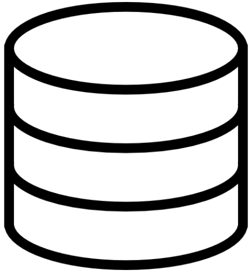


“I think all humans have 3 eyes”

Example: Learning From Data



Peter Washington



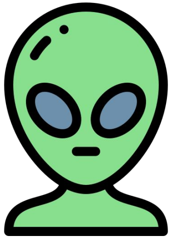
Database

q1

a1

q2

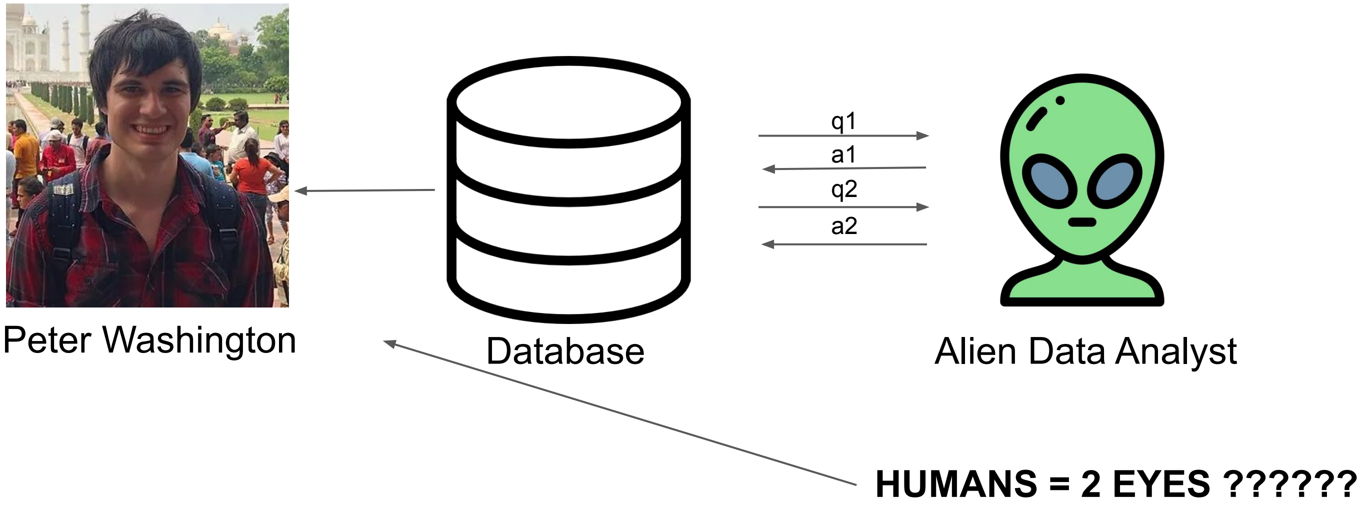
a2



Alien Data Analyst

HUMANS = 2 EYES ??????

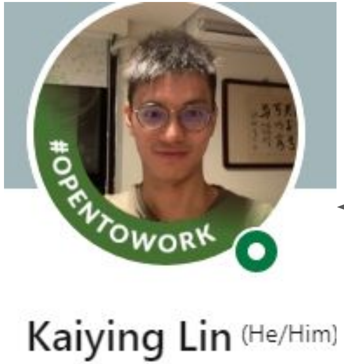
Example: Learning From Data



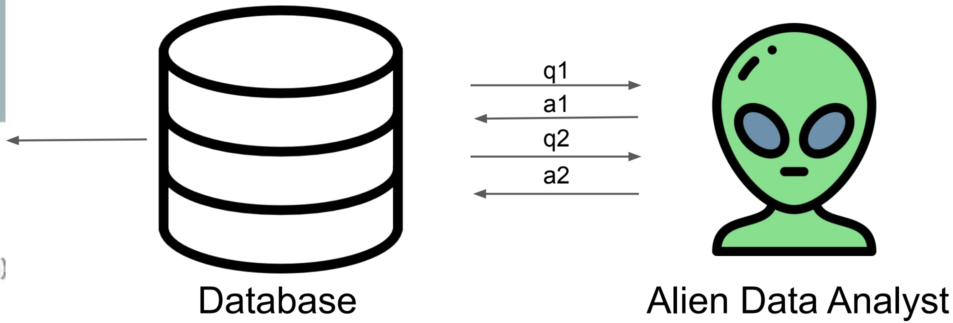
Discussion Question:

Do you think we compromised Peter’s privacy?

Example: Learning From Data



he added me on linkedin
so I used his photo today



- replace Peter Washington with any random member of the population and you will learn the same thing

What is Differential Privacy?

Definition: Differential Privacy

- System for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals in the dataset.
    - Condition: The outcome of any analysis is equally likely, independent of whether any individual joins or refrains from joining, the dataset 🤓

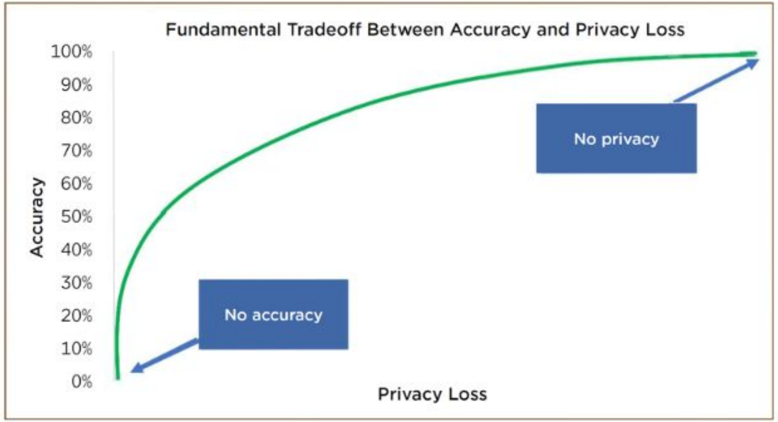
🤓 Definition: Differential Privacy

$$Pr[M(x) \in S] \leq (1 + \in) Pr[M(y) \in S]$$

DP: Privacy-loss Budget

Epsilon (=) Privacy-loss Budget

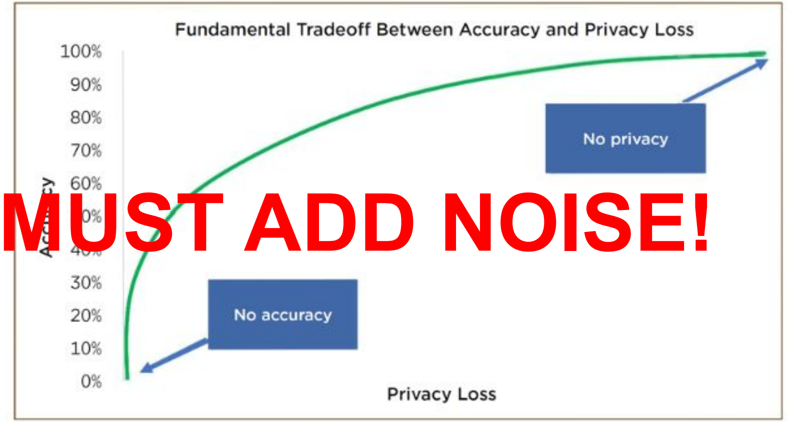
- ε = 0 (perfect privacy), completely useless data
- ε = ∞ (perfect accuracy), completely identifiable data



DP: Privacy-loss Budget

Epsilon (=) Privacy-loss Budget

- $\varepsilon = 0$ (perfect privacy), completely useless data
- $\varepsilon = \infty$ (perfect accuracy), completely identifiable data



WE MUST ADD NOISE!

Privacy-loss Example: Census Bureau

- Prior to 2020, Census did not apply differential privacy to its data
  - Legacy Disclosure Avoidance Methods
    - **Suppression,** suppress values
    - **Coarsening,** round up
    - **Top/bottom coding,** threshold labels ($90,000 or more)
    - **Data swapping,** attributes are swapped
    - **Blank/impute,** attributes are replaced with generated values
    - **Noise injection,** random noise is added to values

Privacy-loss Example: Census Bureau

- Prior to 2020, Census did not apply differential privacy to its data
    - Legacy Disclosure Avoidance Methods
        - **Suppression,** suppress values
        - **Coarsening,** round up
        - **Top/bottom coding,** threshold labels ($90,000 or more)
        - **Data swapping,** attributes are swapped
        - **Blank/impute,** attributes are replaced with generated values
        - **Noise injection,** random noise is added to values

Isn't this enough?

Privacy-loss Example: Census Bureau

- Reconstruction Attack on 2010 Census Bureau
  - reconstructed microdata for 144 million people (46% US population)
  - 76 million reconstructed name, sex, race, ethnicity, with age off by a single year
  - completely re-identify data from 52 million people (17%)

Privacy-loss Example: Census Bureau

- Reconstruction Attack on 2010 Census Bureau
  - reconstructed microdata for 144 million people (46% US population)
  - 76 million reconstructed name, sex, race, ethnicity, with age off by a single year
  - completely re-identify data from 52 million people (17%)

It is not enough!

🤓 Definition: Differential Privacy

$$Pr[M(x) \in S] \leq (1 + \in) Pr[M(y) \in S]$$

Epsilon (=) Privacy-loss Budget

- ε = 0 (perfect privacy), completely useless data
- ε = ∞ (perfect accuracy), completely identifiable data

🤓 Definition: Differential Privacy

$$Pr[M(x) \in S] \leq (1 + \in) Pr[M(y) \in S]$$

Epsilon (=) Privacy-loss Budget

- ε = 0 (perfect privacy), completely useless data
- ε = ∞ (perfect accuracy), completely identifiable data

“If a bad event is very unlikely when I’m not in the dataset (y), then it is still very unlikely when I am (x)”

Characteristics of DP

- Composition
- Group Privacy
- Closure under post-processing

Characteristics of DP: Composition

- joint distribution of the outputs of differentially private mechanisms satisfies differential privacy
  - Sequential composition:
    - if we query ε - different privacy mechanisms t times, and randomization is independent for each query, the then result would be εt- differentially private
  - Parallel composition:
    - If the previous mechanisms are computed on disjoint subsets of the private database then the function g would be the max of εi - differentially private instead

Characteristics of DP: Closure under post-processing

- For any randomized function **F** defined over mechanism **M, if M** satisfies ε differential privacy, so does **F(M)**
- (Composition + Post-processing) = 👍 (PLB)

Characteristics of DP: Group Privacy

- ε-differential privacy protects databases which differ in one row
  - extend to protect databases which differs in c rows
- allows the control of privacy loss acquired by groups

Private mechanisms DP

- Sensitivity
- The Laplace mechanism
- Randomized response
- Stable Transformations

Private mechanisms DP

- Sensitivity
- The Laplace mechanism
- Randomized response ← **This has less math!**
- Stable Transformations

Sensitivity (1)

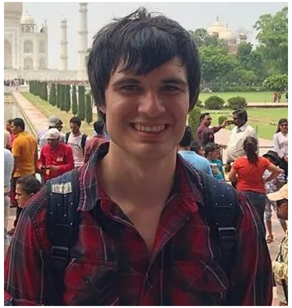
- Impact a change in the underlying data set can have on the result of the query
    - the maximum possib

$$Sensitivity = \max_{x_A, x_B \subseteq X} \|q(x_A) - q(x_B)\|_1$$

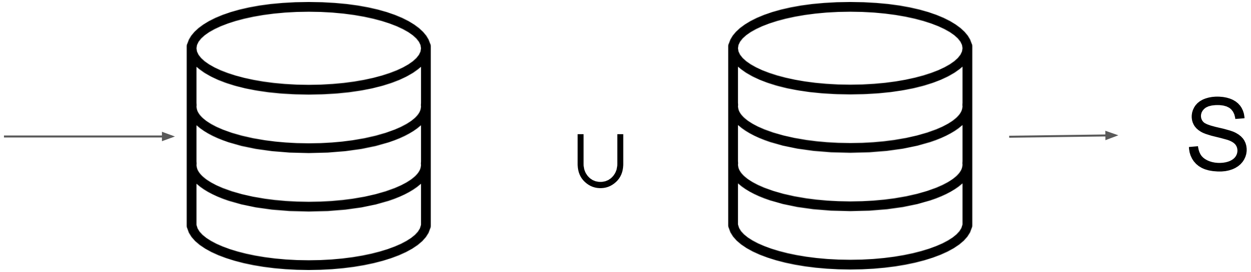
Sensitivity (1)

- Impact a change in the underlying data set can have on the result of the query
  - the maximum possib

$$Sensitivity = \max_{x_A, x_B \subseteq X} \|q(x_A) - q(x_B)\|_1$$



Peter Washington



Database V2

Database V1

Laplace Mechanism (2)

$$F(x) = f(x) + \mathsf{Lap}\left(\frac{s}{\epsilon}\right)$$

- how much noise, and what kind of noise?
- symmetric version of exponential distribution
  - f(x) = some function
  - Lap(S) = sampling from L.D. with center 0 and scale S
  - s is the sensitivity

Laplace Mechanism (2)

```
adult [adult ['Age'] >= 40].shape[0]

//Returns 14,237 people
```

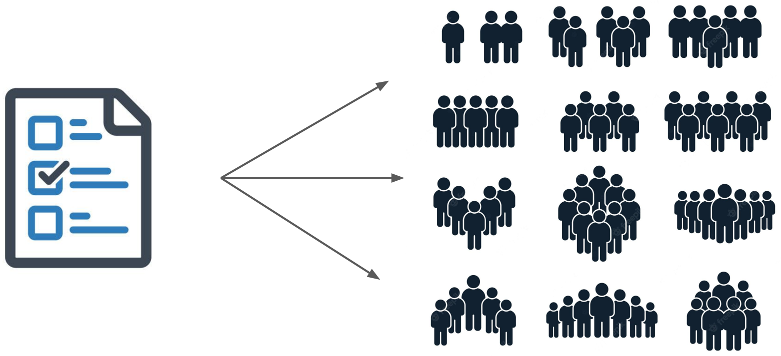
Laplace Mechanism (2)

```
adult [adult ['Age'] >= 40].shape[0]

//Returns 14,237 people
```

```
sensitivity = 1
epsilon = 0.1
adult [adult ['Age'] >= 40].shape[0] +
   np.random.laplace(loc=0, scale=sensitivity/epsilon)

//Returns 14240.232560364662 people
```

Randomized Response (3)

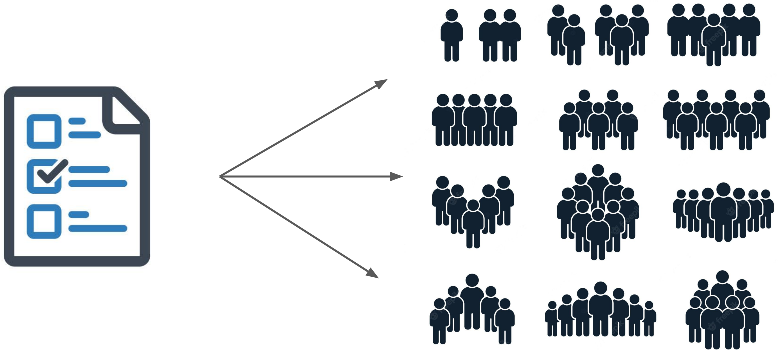
Randomized Response (3)



Survey

People

Randomized Response (3)



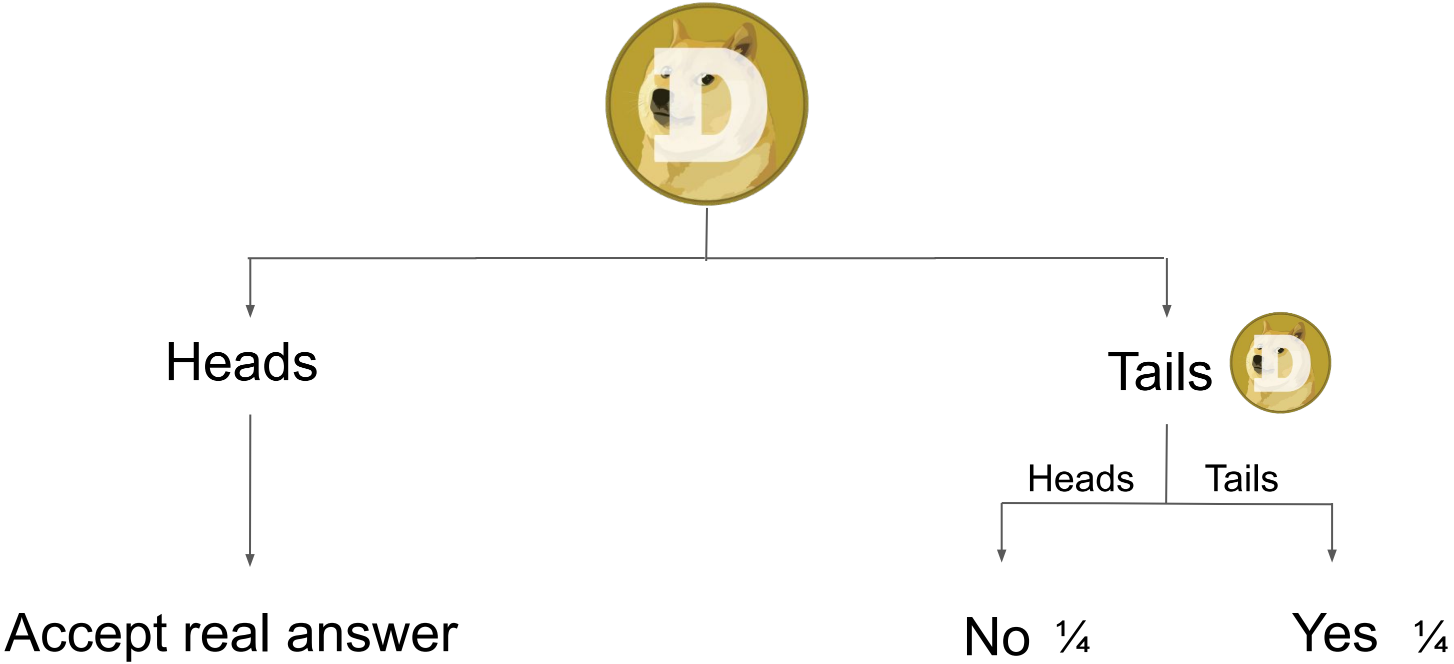
Survey

People

“Do you pick your nose?”

Randomized Response (3)

Randomized Response (3)



Stable Transformations (4)

- transformations applied to a dataset that allows differential privacy

$$|T(A) \oplus T(B)| \leq c \times |A \oplus B|$$

- T is c-stable if for any two input data sets A and B
  - c*ε-differential privacy

Part 2: Discussion Paper

Part 2: Discussion Paper

https://tinyurl.com/4d5ezxfe

https://aboutmyinfo.org/identity/samples

Part 2: Discussion Paper

https://tinyurl.com/4d5ezxfe

https://aboutmyinfo.org/identity/samples

AGE, GENDER, ZIP CODE

The Paper: 💀

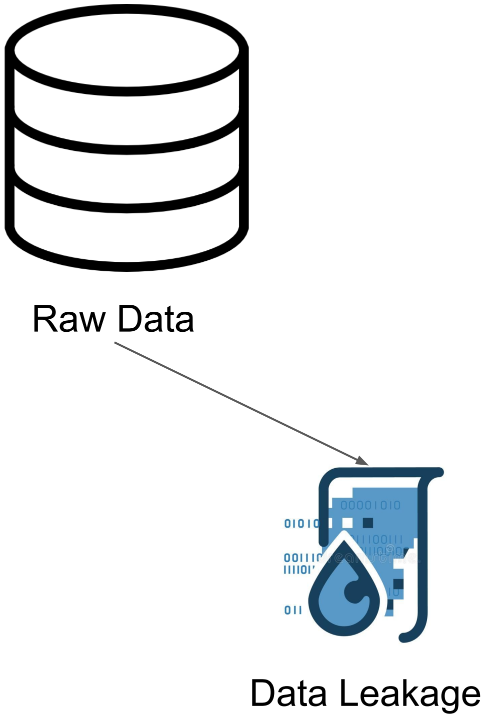
- Title:
  - Differential Privacy Protection Against Membership Inference Attack on Machine Learning for Genomic Data
- Authors:
  - Junjie Chen, Wendy Hui Wang and Xinghua Shi
- Problem:
  - Genome privacy is a growing concern in machine learning

The Paper: 💀

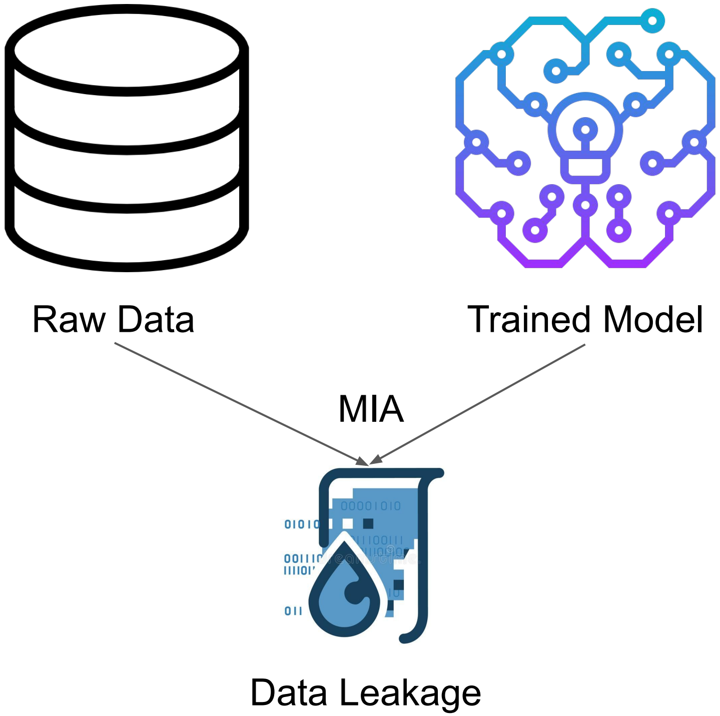
- Title:
  - Differential Privacy Protection Against Membership Inference Attack on Machine Learning for Genomic Data
- Authors:
  - Junjie Chen, Wendy Hui Wang and Xinghua Shi
- Problem:
  - Genome privacy is a growing concern in machine learning

why?

The Paper: 💀



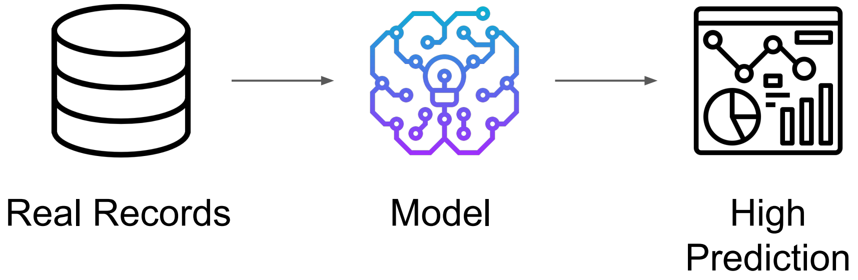
The Paper: 💀



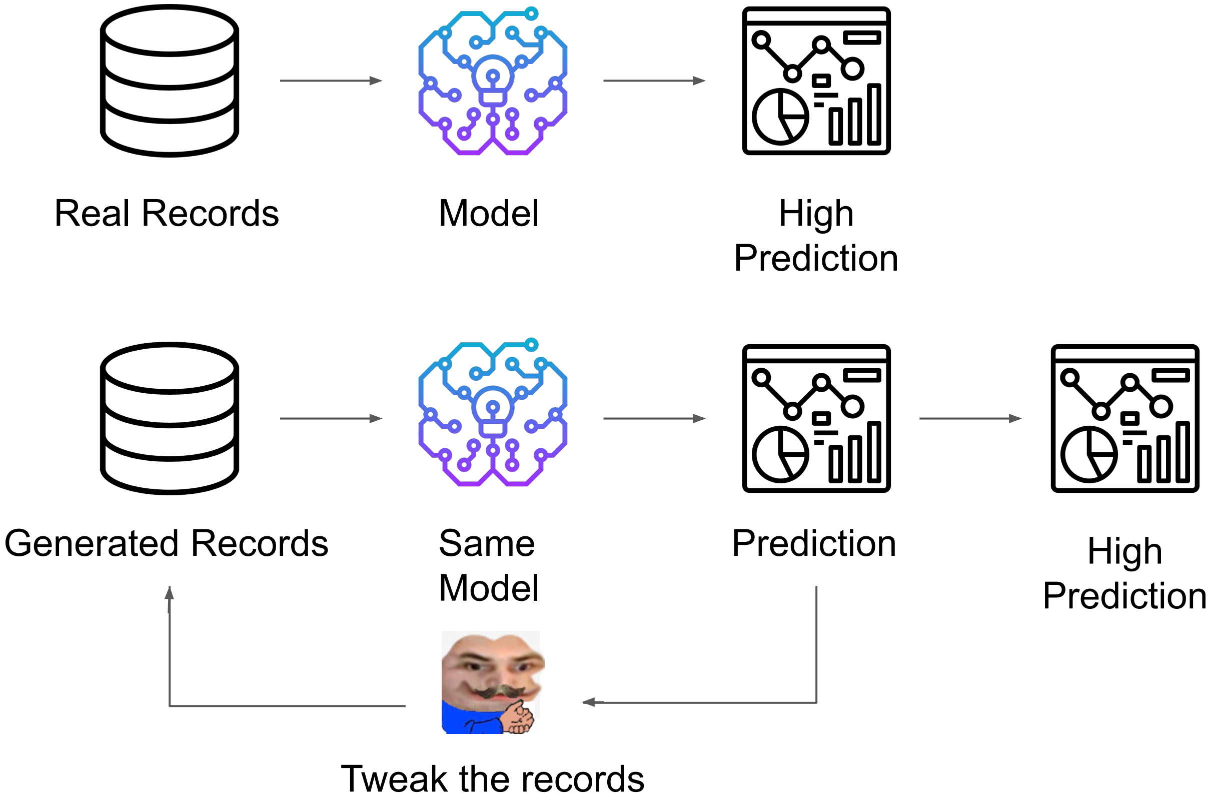
The Paper: 💀

- Split yeast data 50-50,
  - ½ private target dataset, ½ public shadow dataset
  - Split public shadow dataset again 80-20
  - 80% model training, 20% for ground truth
- White-box model atack
  - worst case privacy leak
- 2 ML models
  - Lasso
  - CNN

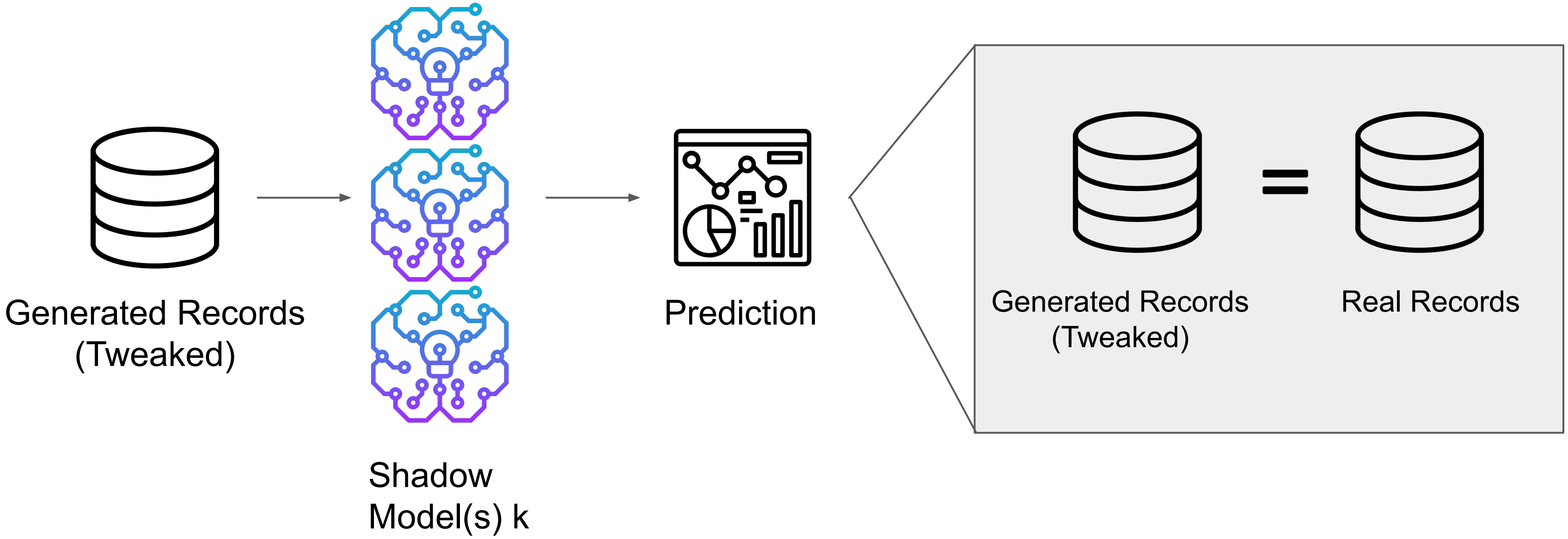
Membership Interference Attack



Membership Interference Attack



Membership Interference Attack



The Paper: Findings

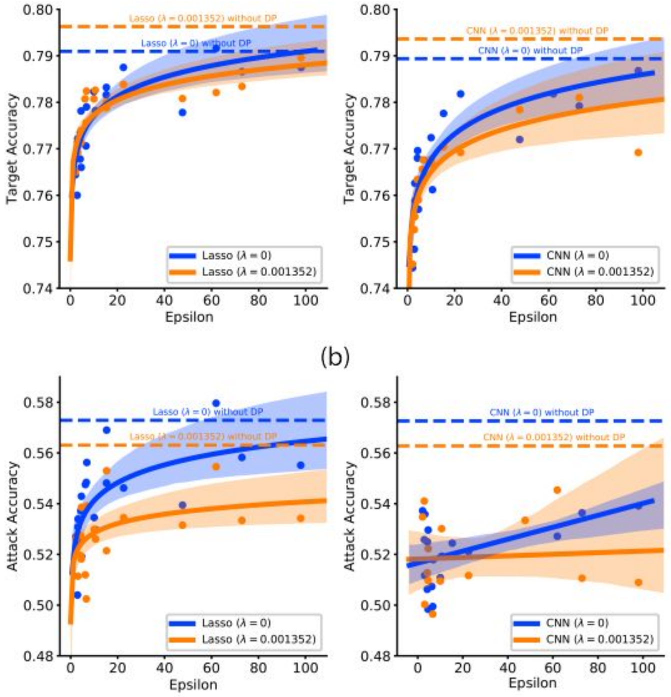
- The attack accuracy of MIA on Lasso and CNN with no sparsity
  - 0.5728, 0.5726 respectively,

Table 1. **Model performance against MIA (without DP).**

Methods	Target model Accuracy	Target model Std.	Attack model Accuracy	Attack model Std.
Lasso ($\lambda = 0$)	0.7910	0.0123	0.5728	0.0071
Lasso ($\lambda = 0.001352$)	0.7963	0.0157	0.5631	0.0042
CNN ($\lambda = 0$)	0.7894	0.0199	0.5726	0.0059
CNN ($\lambda = 0.001352$)	0.7936	0.0225	0.5628	0.0050

The Paper: Findings

- There exists a trade-off between privacy and accuracy of target models
- Lasso:
    - A smaller privacy budget (ε ≤ 10)
        - rapidly reduces attack accuracy
    - A bigger privacy budget (ε > 10)
        - attack accuracy stays relatively stable
- CNN:
    - Attack accuracy decreases when ε increases



The Paper: Findings

- Model Sparsity
  - model sparsity can improve the accuracy of the target model and reduce the attack accuracy of MIA when DP is not deployed
  - sparse models have slightly worse target model accuracy under different privacy budgets
    - ε < 10: privacy budget is smaller than the trade-off
    - ε > 10: accuracy of target model is incentive to model sparsity with larger privacy budgets
  - sparse models provide better privacy protection than without sparsity, given the same DP budget

Table 1. **Model performance against MIA (without DP).**

Methods	Target model Accuracy	Target model Std.	Attack model Accuracy	Attack model Std.
Lasso ($\lambda = 0$)	0.7910	0.0123	0.5728	0.0071
Lasso ($\lambda = 0.001352$)	0.7963	0.0157	0.5631	0.0042
CNN ($\lambda = 0$)	0.7894	0.0199	0.5726	0.0059
CNN ($\lambda = 0.001352$)	0.7936	0.0225	0.5628	0.0050

The Paper: TLDR

- Data leakage is possible from datasets and training models.
- Determine a good balance between data privacy and prediction accuracy based on the privacy loss budget.
- It doesn’t hurt to sparse your data.

Discussion Questions!

1. Have you ever used differential privacy (or similar methods) in your own research? What did you do with your data to maintain privacy?
2. Why do you think there isn't a legal mandate for differential privacy across all "official" databases? (i.e., Census started in 2020 💀)
3. Do you truly have a freedom of choice when it comes to opting-in or out of a database? (de-identification problem)
4. To what extent is the organization that holds the data liable for data leakage? Or are they not liable enough?
5. Do individuals have the right to their own data? (i.e., PHI)
6. Can you think of another example where you could use a membership interference attack? (i.e., phenotype prediction, genomic data)