

# Privacy-preserving machine learning and differential privacy

Antti Honkela

Finnish Center for Artificial Intelligence FCAI,  
Department of Mathematics and Statistics &  
Department of Public Health,  
University of Helsinki

AI Day, 12 December 2018

**FCAI** Finnish  
Center for  
Artificial  
Intelligence

# SUN ON PRIVACY: 'GET OVER IT'

THE CHIEF EXECUTIVE officer of Sun Microsystems said Monday that consumer privacy issues are a "red herring."

"You have zero privacy anyway," Scott McNealy told a group of reporters and analysts Monday night at an event to launch his company's new Jini technology.

"Get over it."



About Issues Our Work Take Action

## Google CEO Eric Schmidt Dismisses the Importance of Privacy

NEWS UPDATE BY RICHARD ESGUERRA | DECEMBER 10, 2009

Yesterday, the web was buzzing with commentary about Google CEO Eric Schmidt's dangerous, dismissive response to concerns about search engine users' privacy. When asked during an interview for CNBC's recent "Inside the Mind of Google" special about whether users should be sharing information with Google as if it were a "trusted friend," Schmidt responded, "If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place."



## Google CEO: Secrets Are for Filthy People



Ryan Tate

270.56K

12/04/09 04:48PM Filed to: GOOGLEPLEX

### Privacy is dead on Facebook. Get over it.

Cool kids don't care about privacy, claim CEOs. So, neither should you.

Recommend 0

Tweet

G+

Share 93

Below: Discuss Related



THINK



By Helen A.S. Popkin

msnbc.com updated 1/13/2010 8:56:58 AM ET

Print | Font: A + -

Once upon a time at Facebook, or so the story from an anonymous Facebook employee goes, there was a general password employees could use to access Facebook accounts. For kicks and giggles, some Facebook employees, including the one recently interviewed on the [Rumpus Web site](#), did just that.

Two Facebook employees got fired, says Anonymous Facebook Employee, for manipulating user profile information. Others, such as Anonymous Facebook Employee, just peeked.



# WHY 'ANONYMOUS' DATA SOMETIMES ISN'T

LAST YEAR, NETFLIX published 10 million movie rankings by 500,000 customers, as part of a challenge for people to come up with better recommendation systems than the one the company was using. The data was anonymized by removing personal details and replacing names with random numbers, to protect the privacy of the recommenders.

Arvind Narayanan and Vitaly Shmatikov, researchers at the University of Texas at Austin, [de-anonymized some of the Netflix data](#) by comparing rankings and timestamps with public information in the [Internet Movie Database](#), or IMDb.

## Identifying Personal Genomes by Surname Inference

Melissa Gymrek<sup>1,2,3,4</sup>, Amy L. McGuire<sup>5</sup>, David Golan<sup>6</sup>, Eran Halperin<sup>7,8,9</sup>, Yaniv Erlich<sup>1,\*</sup>

+ See all authors and affiliations

Science 18 Jan 2013;  
Vol. 339, Issue 6117, pp. 321-324  
DOI: 10.1126/science.1229566

Article

Figures & Data

Info & Metrics

eLetters



Abstract

# A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

## “Anonymized” data really isn't—and here's why not

Companies continue to store and sometimes release vast databases of “ ...

NATE ANDERSON · 9/8/2009, 2:25 PM

41



The Massachusetts Group Insurance Commission had a bright idea back in the mid-1990s—it decided to release “anonymized” data on state employees that showed every single hospital visit. The goal was to help researchers, and the state spent time removing all obvious identifiers such as name, address, and Social Security number. But a graduate student in computer science saw a chance to make a point about the limits of anonymization.

Latanya Sweeney requested a copy of the data and went to work on her “reidentification” quest. It didn't prove difficult. Law professor Paul Ohm describes Sweeney's work:

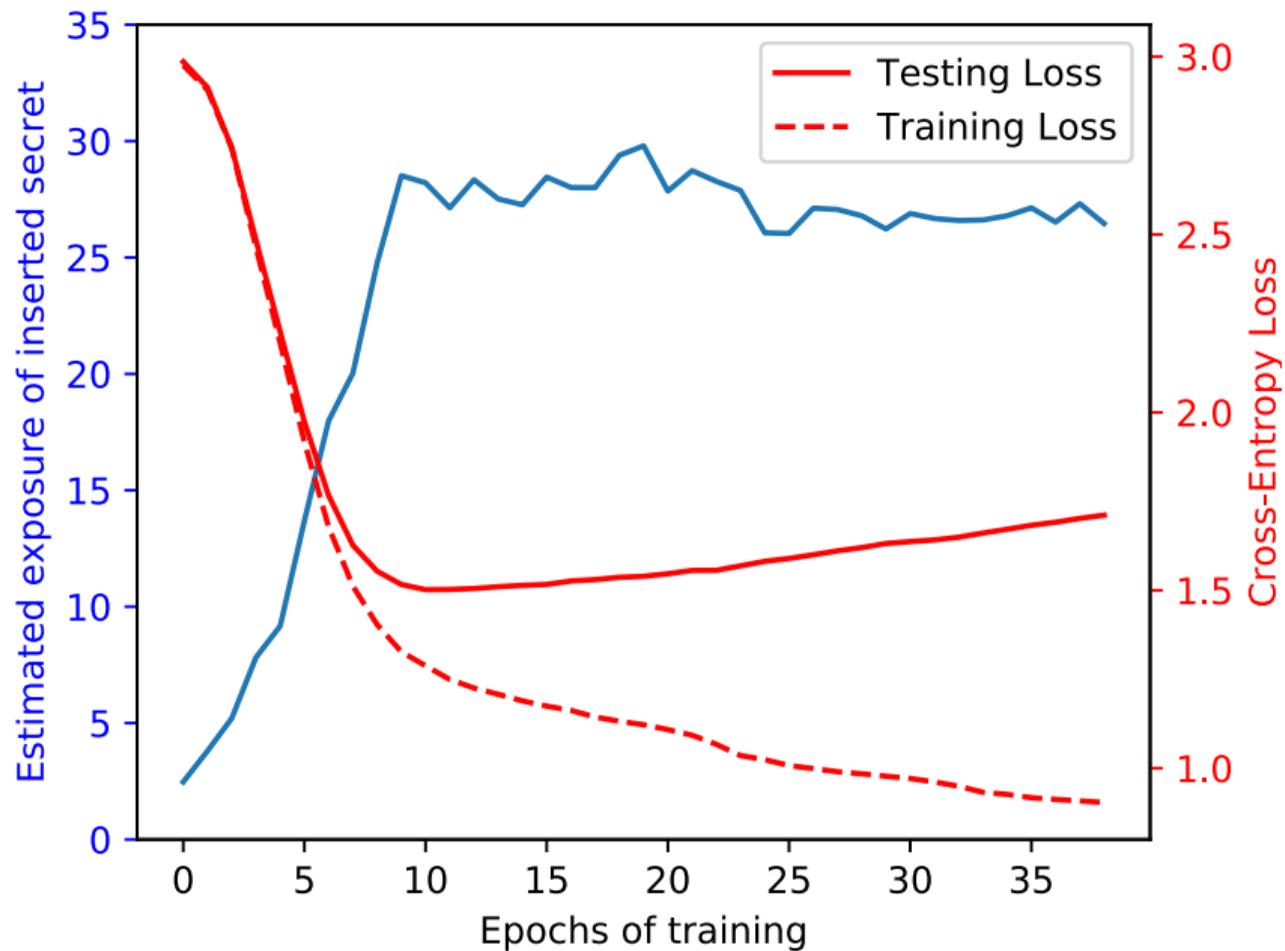
“

At the time GIC released the data, William Weld, then Governor of Massachusetts, assured the public that GIC had protected patient privacy by deleting identifiers. In response, then-graduate student Sweeney started hunting for the Governor's hospital records in the GIC data. She knew that Governor Weld resided in Cambridge, Massachusetts, a city of 54,000 residents and seven ZIP codes. For twenty dollars, she purchased the complete voter rolls from the city of Cambridge, a database containing,

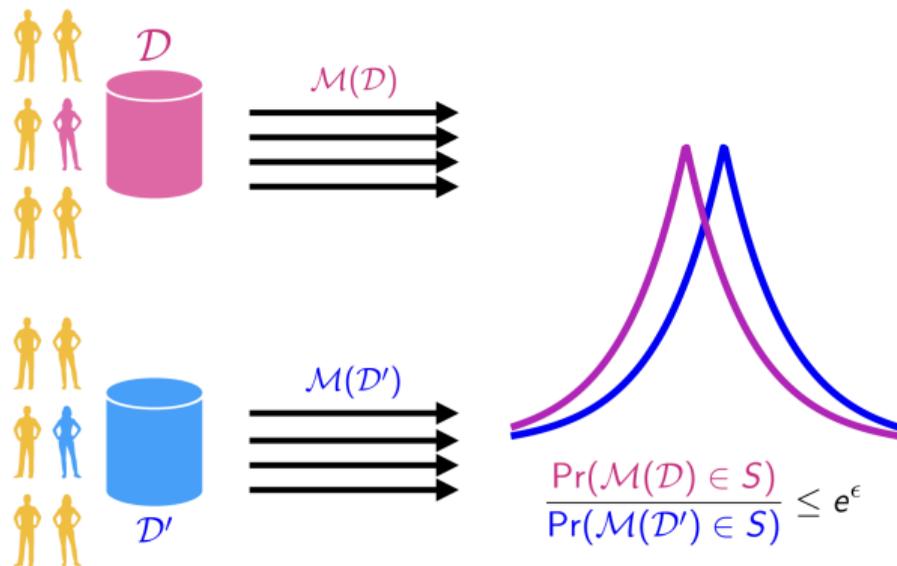


**Figure 1: An image recovered using a new model inversion attack (left) and a training set image of the victim (right). The attacker is given only the person's name and access to a facial recognition system that returns a class confidence score.**





# Differential privacy (DP; Dwork *et al.*, 2006)



- ▶ Provides protection against adversaries with side information
- ▶ Is invariant to post-processing
- ▶ Degrades gracefully under composition

## Example: Randomised response

Assume respondents are instructed to answer a potentially embarrassing query as follows:

1. Flip a coin.
2. If **tails**, then respond truthfully.
3. If **heads**, then flip a second coin and respond “Yes” if heads and “No” if tails.

## Example: Randomised response

Assume respondents are instructed to answer a potentially embarrassing query as follows:

1. Flip a coin.
2. If **tails**, then respond truthfully.
3. If **heads**, the flip a second coin and respond “Yes” if heads and “No” if tails.

This mechanism is  $\epsilon$ -DP with  $\epsilon = \ln 3$ .

Proof.

Analysis of the cases shows 3/4 probability to answer truthfully.

$$\frac{\Pr(\text{Yes} \mid \text{Yes})}{\Pr(\text{Yes} \mid \text{No})} = \frac{3/4}{1/4} = \frac{\Pr(\text{No} \mid \text{No})}{\Pr(\text{No} \mid \text{Yes})} = 3.$$

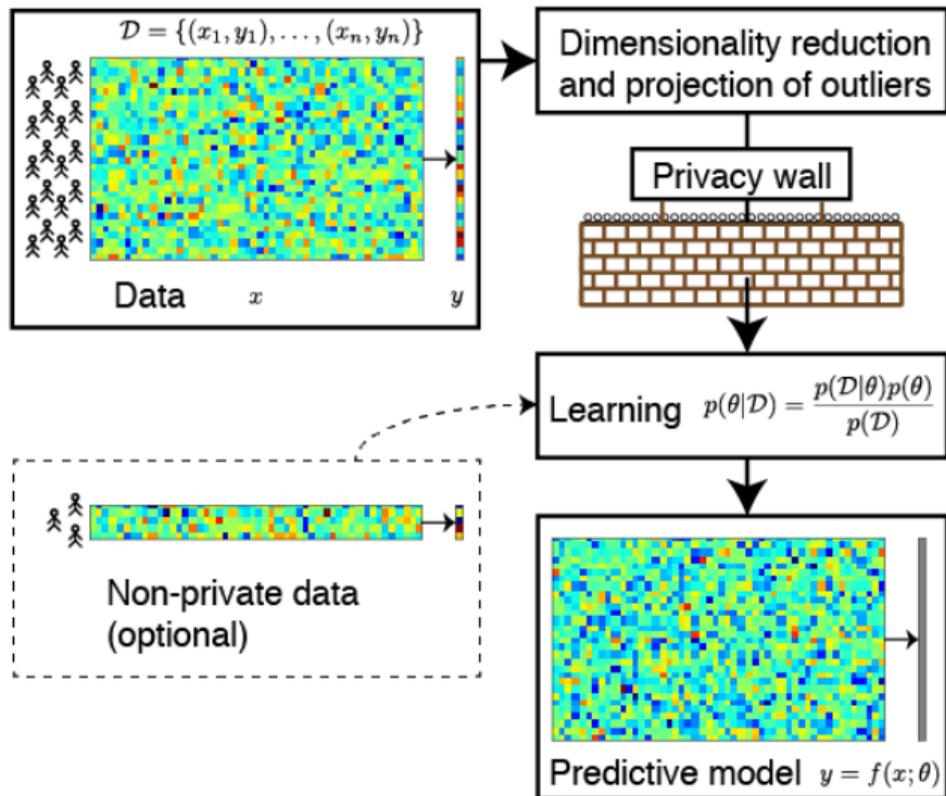


## Practical DP algorithm: DP stochastic gradient descent

Assume objective  $\mathcal{F}(\theta, X) = \sum_i \mathcal{F}_i(\theta, x_i)$  depending on data set  $X = (x_1, \dots, x_n)$ , where each sample comes from a different individual whose privacy we wish to protect

1. Each  $g_i(\theta) = \nabla_{\theta} \mathcal{F}_i(\theta, x_i)$  is clipped s.t.  $\|g_i(\theta)\|_2 \leq c_t$  in order to calculate *gradient sensitivity*
2. Subsampling  $x_i$  with frequency  $q$  provides *privacy amplification* from subsampling
3. Gradient contributions from all data samples in the mini batch are summed and perturbed with Gaussian noise  $\mathcal{N}(0, \sigma^2 \mathbf{I})$
4. Total privacy cost can be computed from composition theorems or using the *moments accountant* (Abadi *et al.*, CCS 2016)

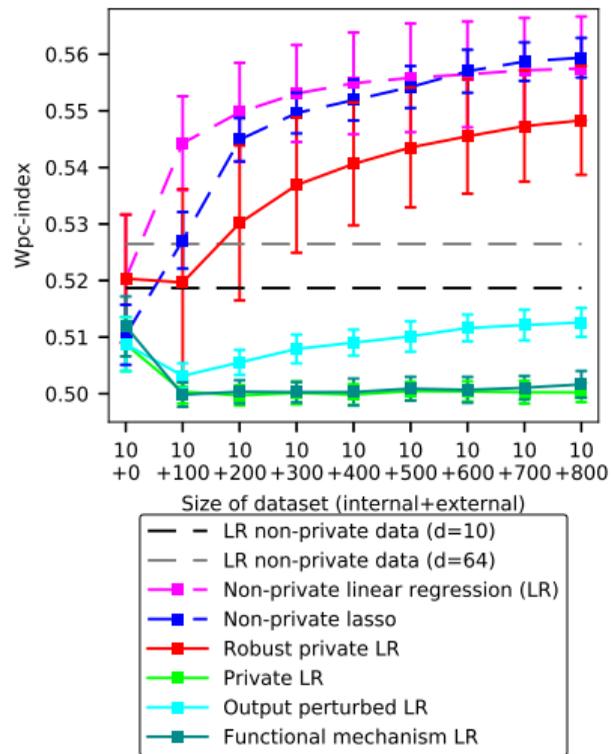
# DP machine learning



# DP machine learning applications

- ▶ DP versions of most common ML algorithms
  - ▶ Linear and logistic regression
  - ▶ Mixture models and clustering
  - ▶ Deep neural networks
- ▶ Example: predicting cancer drug efficacy using gene expression
  - ▶ 800 cell lines, averaging accuracy over 124 drugs
  - ▶ Method: linear regression
  - ▶ Dimensionality reduction using prior knowledge on most important cancer genes

# DP linear regression for drug sensitivity prediction



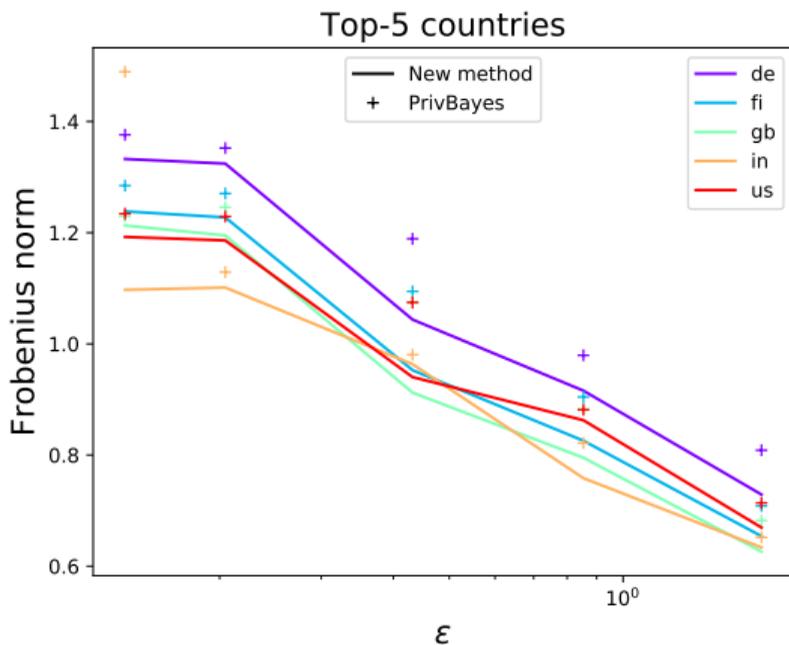
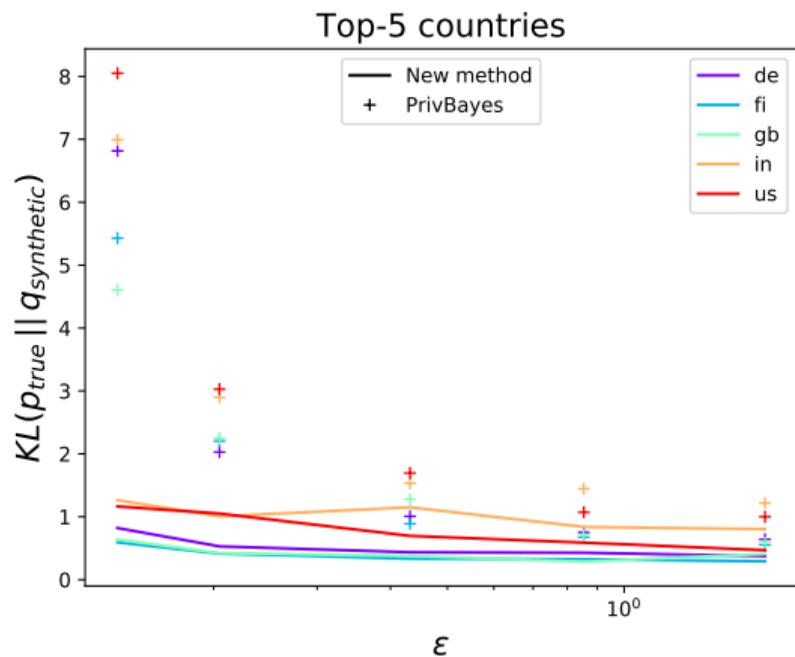
## Challenges with DP learning

- ▶ High dimensionality makes DP learning more difficult
  - ▶ Aggressive dimensionality reduction necessary
- ▶ DP guarantee is worst case over all possible data sets
  - ▶ Eliminating outliers can help a lot
- ▶ Learning complex tasks from scratch is very hard
  - ▶ Using additional non-private can be very helpful

## DP data release

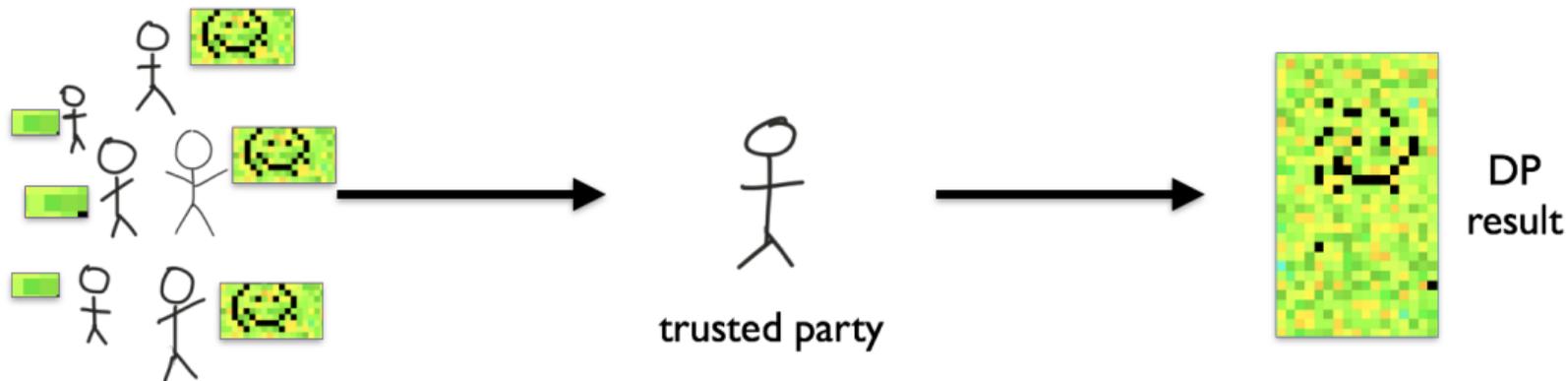
- ▶ Important use for privacy-preserving ML: releasing an anonymised version of a data set
- ▶ Generative modelling approach:  
Data  $\rightarrow$  Generative model  $\rightarrow$  Generated data
- ▶ Training the model under DP guarantees the data will be DP
- ▶ Effectively: we will have a synthetic data set with similar statistical properties as the original, but with no identifiable entries

# DP data release for mobile app usage data



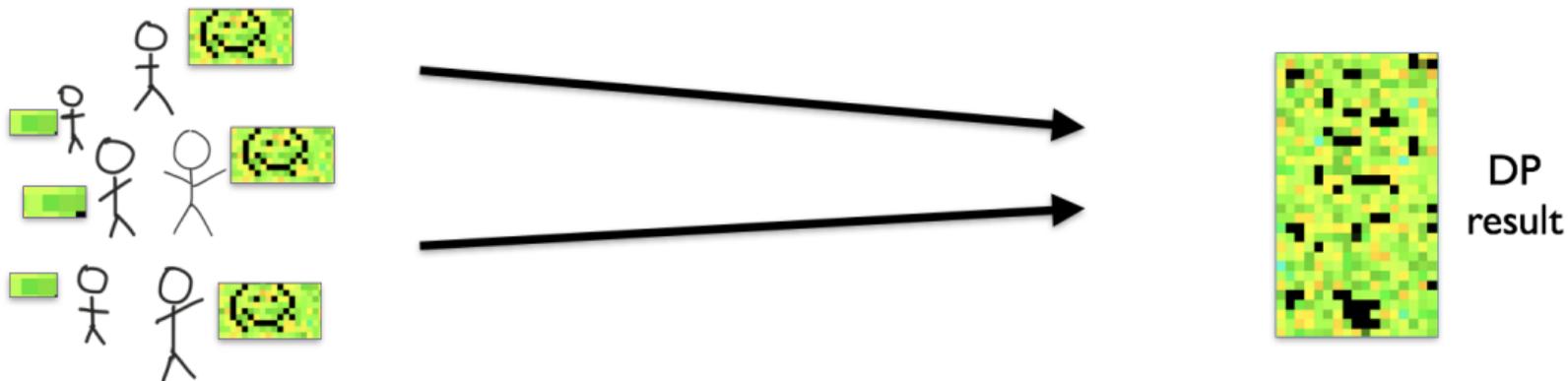
## DP learning with distributed data

- ▶ DP is necessary to ensure the model does not leak private information, but does not protect the learning process
- ▶ Combining with cryptography allows efficient secure and private learning with distributed data



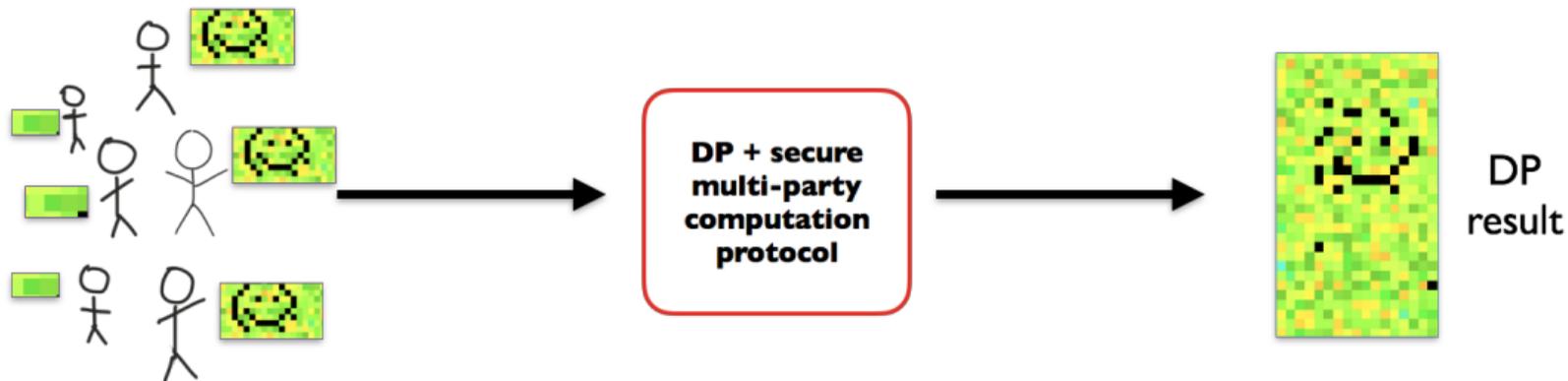
## DP learning with distributed data

- ▶ DP is necessary to ensure the model does not leak private information, but does not protect the learning process
- ▶ Combining with cryptography allows efficient secure and private learning with distributed data



## DP learning with distributed data

- ▶ DP is necessary to ensure the model does not leak private information, but does not protect the learning process
- ▶ Combining with cryptography allows efficient secure and private learning with distributed data



## Conclusion

- ▶ ML models remember their training data, can compromise privacy of training data subjects
- ▶ Differential privacy (DP) can provide strong privacy guarantees, but may limit the accuracy especially for more complex tasks
- ▶ Effective DP learning requires a different approach from standard ML: dimensionality reduction, robustness

# Acknowledgements

Mrinal Das

Onur Dikmen

Mikko Heikkilä

Joonas Jälkö

Antti Koskela

Eemil Lagerspetz

Arttu Nieminen

Teppo Niinimäki

Sasu Tarkoma

Kana Shimizu

Samuel Kaski

Funding: Academy of Finland