PRIVACY PRESERVING MACHINE LEARNING

or

How to do AI on Sensitive Data

INDEX

- Some Words on Privacy
- Why Pseudonymization is Not the Solution
- Homomorphic Encryption
- Federated Learning

Some Words on Privacy

WHY PRIVACY?

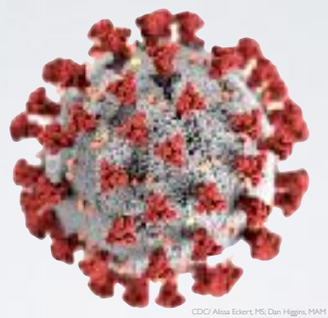
- Ethics / Moral
- Legal
- Data Quality

Chris Whippet / One Nation under CCTV

There are three good reasons for data privacy

Some Words on Privacy

- >10M cases



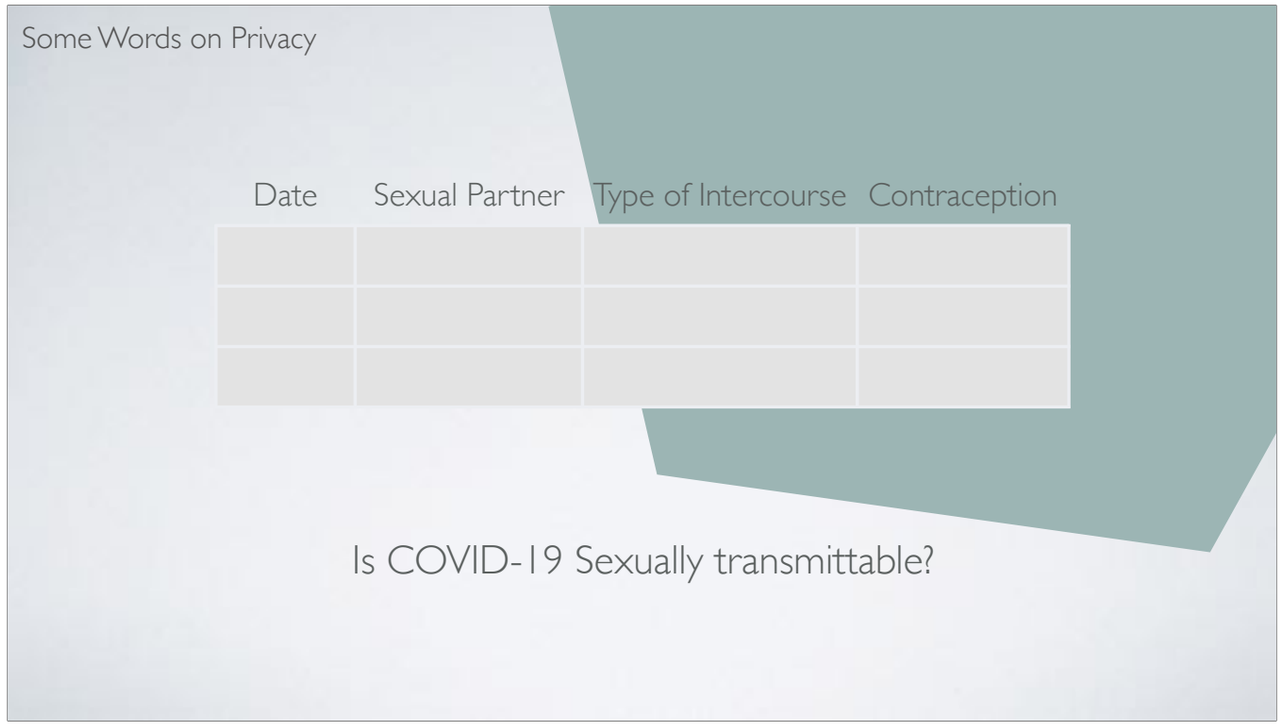
CDC/ Alissa Eckert, MS; Dan Higgins, MAM

COVID-19

Early evidence for:

- Masks may work
- Children may not be as infectious
- Blood group A may have increased Risk

One might expect more results from 10 million data-points but several problems hinder data acquisition one of which is privacy



This example illustrates all three reasons why data privacy is important.

WHY PSEUDONYMIZATION IS NOT THE SOLUTION

Injection Attacks and Differential Privacy

Why Pseudonymization is Not the Solution

INJECTION ATTACKS

Is Dr. Max Musterman a murderer?

Sex	PhD	Murderer
Male	No	No
Female	No	No
Male	Yes	Yes
Female	Yes	No

The dataset reveals Max’s crime.

Why Pseudonymization is Not the Solution

THE NUMBERS GAME

Population of Hamburg $1.787M$

Age 80 choices

Sex 3 Choices

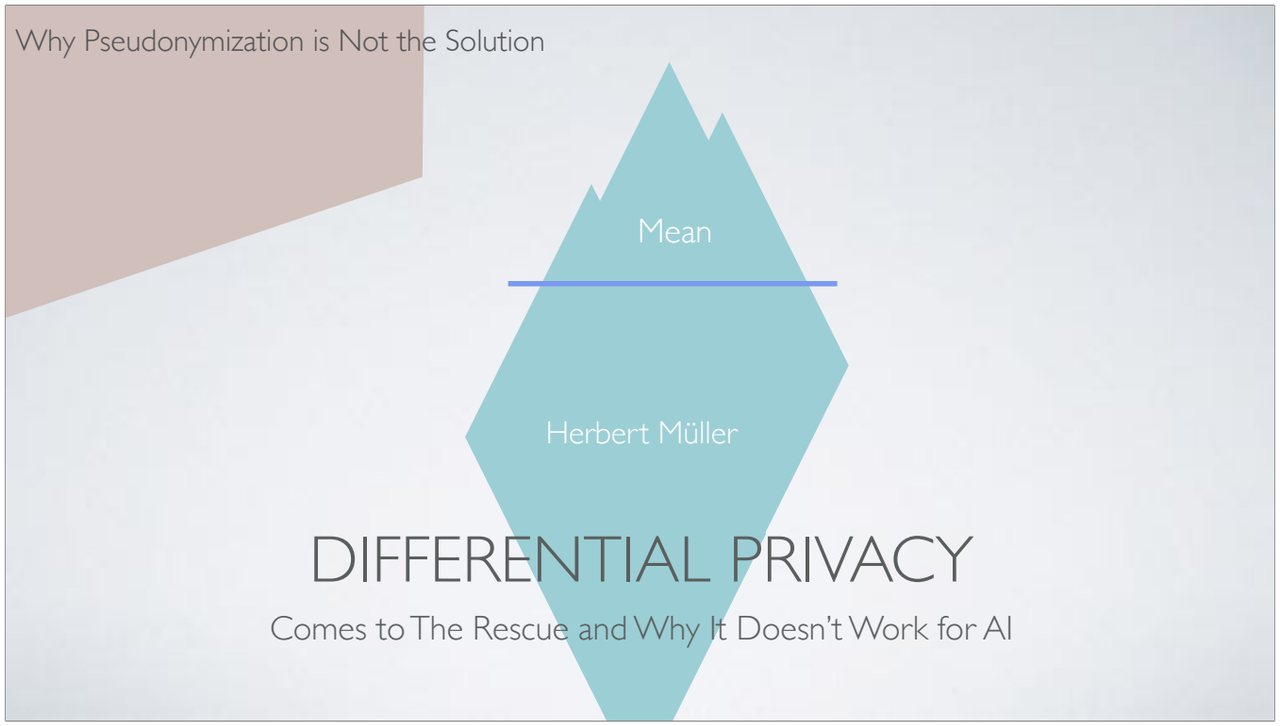
104 City districts

50 Choices for Body height

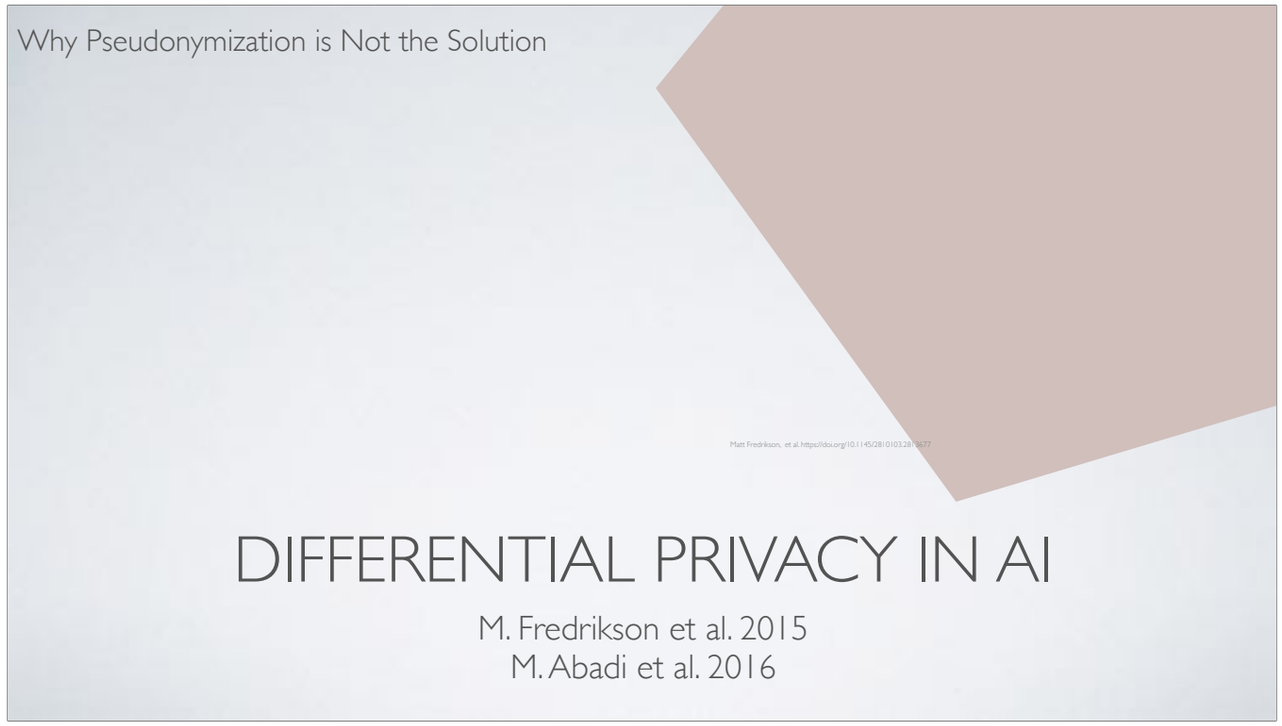
$80 \cdot 3 \cdot 104 \cdot 50 = 1.248M$

About one combination per inhabitant

The number of choices for each guessable entry (quasi-identifier) multiply resulting in a combinatorial explosion.



Differential privacy limits access to the dataset to provide plausible deniability.
However, the limited data harms AI training.



Differential privacy can be used to prevent a model inversion attack where the attacker tries to recover the data used to train a Model.

HOMOMORPHIC ENCRYPTION

or

How to Compute on encrypted Data

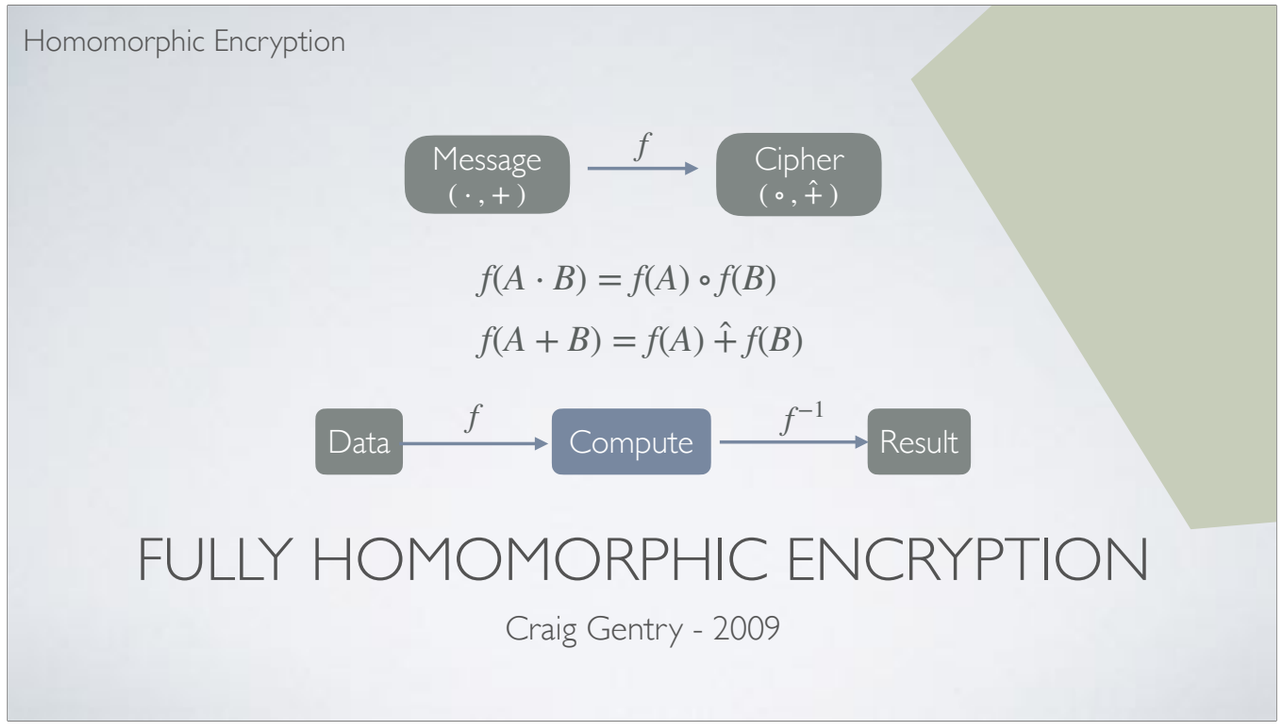
Homomorphic Encryption

SUBSTITUTION CIPHERS

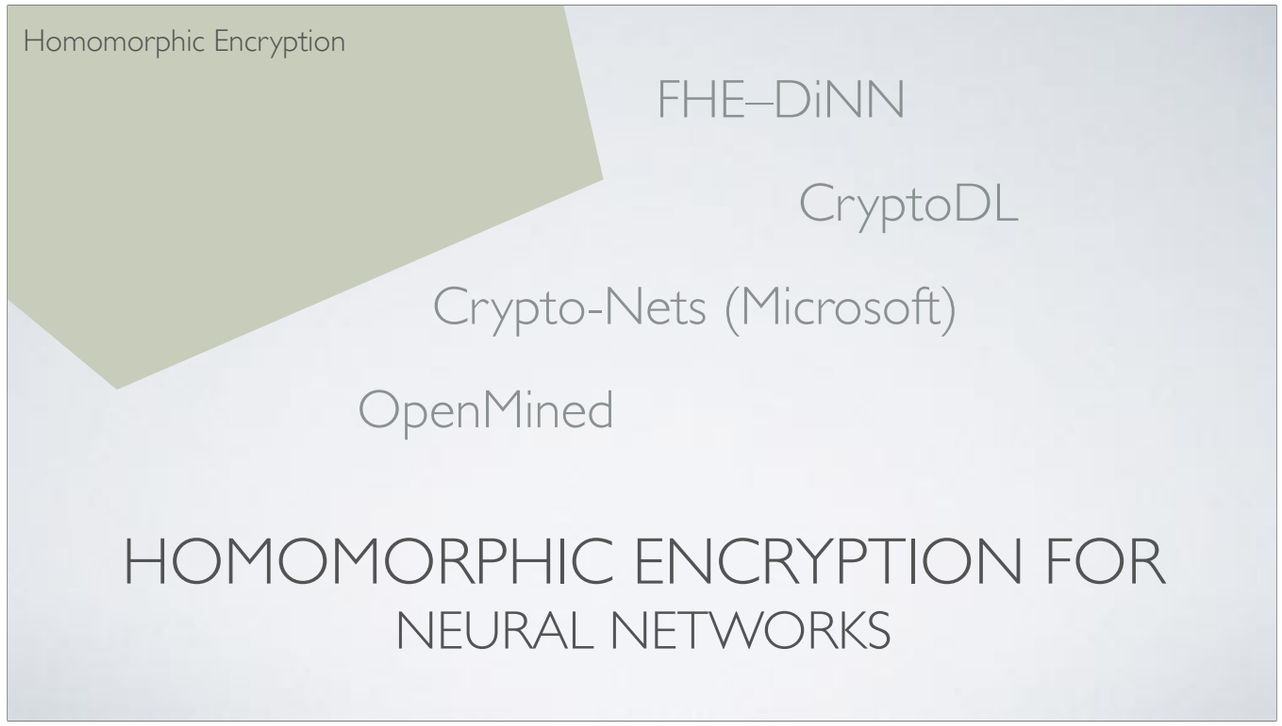
KP VJG DGIKPPKPI VJGTG
YCU IQF CPF VJGP JG
OCFG OCP

- Can you read this?
- But can you replace " IQF" with "OCP" and "OCP" with "TQDQVU"

This example illustrates how some encryptions allow operations on the encrypted data.



Homomorphic encryption allows multiplications and additions of the message without knowledge of the key and thus the message.



There are several projects implementing Homomorphic Encryption for neural networks.

Homomorphic Encryption

“there is a world market for maybe five computers”

- Thomas J. Watson

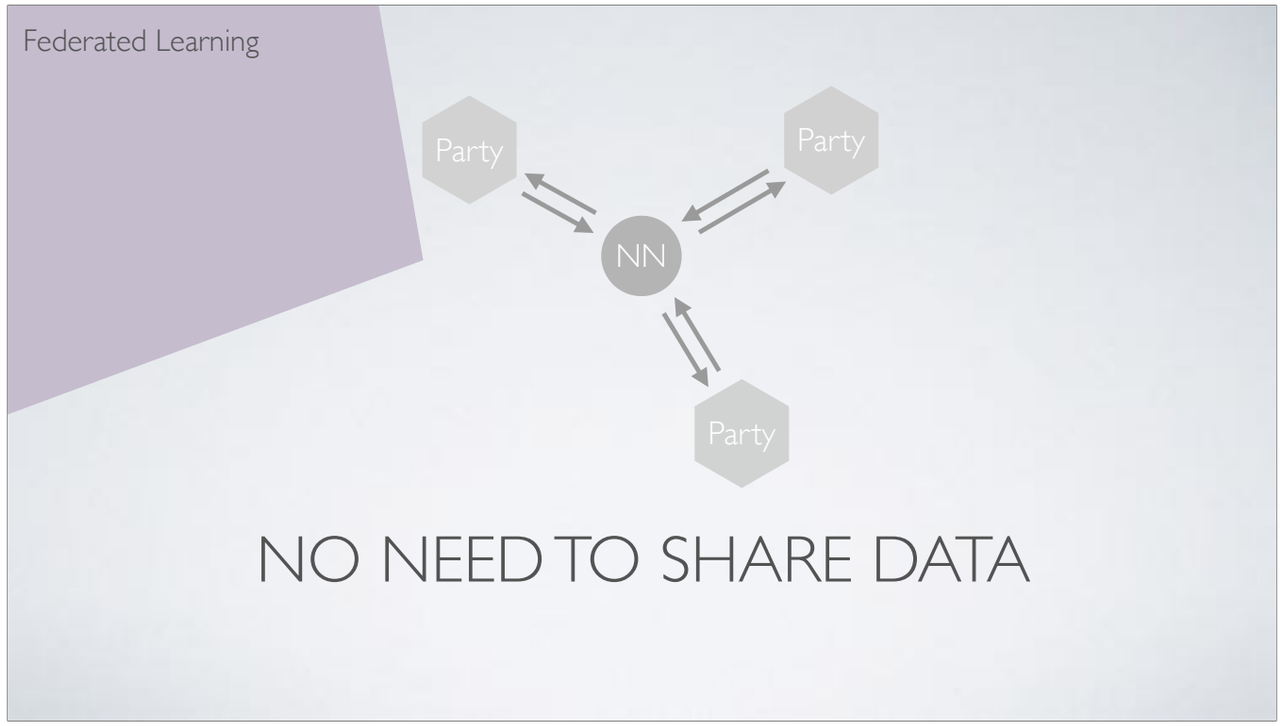
WHY IS A PHYSICIST INTERESTED IN SECURE MULTIPARTY COMPUTATION?

Quantum computes are expected to remain in few numbers for the foreseeable time.
Quantum Homomorphic Encryption was developed to allow quantum computation of sensitive data.

FEDERATED LEARNING

Or

How To Collude in AI without Trust



In federated learning, a neural network is trained by multiple parties without sharing data.

"I MAY NOT HAVE GONE WHERE INTENDED TO GO,
BUT I THINK I HAVE ENDED UP
WHERE I NEEDED TO BE"

- DOUGLAS ADAMS

Thank You for Listening

Jakob Teuffel

j.teuffel@icloud.com

Linkedin

David Altrath