Federated Learning and Privacy-Preserving Machine Learning

Imperial College London

Jonathan Passerat-Palmbach, Anna Beer

Equideum Health | A ConsenSys Mesh Portfolio Company

CONSENSYS HEALTH

Why **Private, Secure & Verifiable** ML?

From the Headlines


Apple contractors 'regularly hear confidential details' on Siri recordings

Workers hear drug deals, medical details and people having sex, says whistleblower


Alexa has been eavesdropping on you this whole time

When Alexa runs your home, Amazon tracks you in more ways than you might want.


Facebook Contractors Have Been Listening to 'Hey Portal'

Bloomberg **Kurt Wagner and Mark Gurman**

Bloomberg September 18, 2019

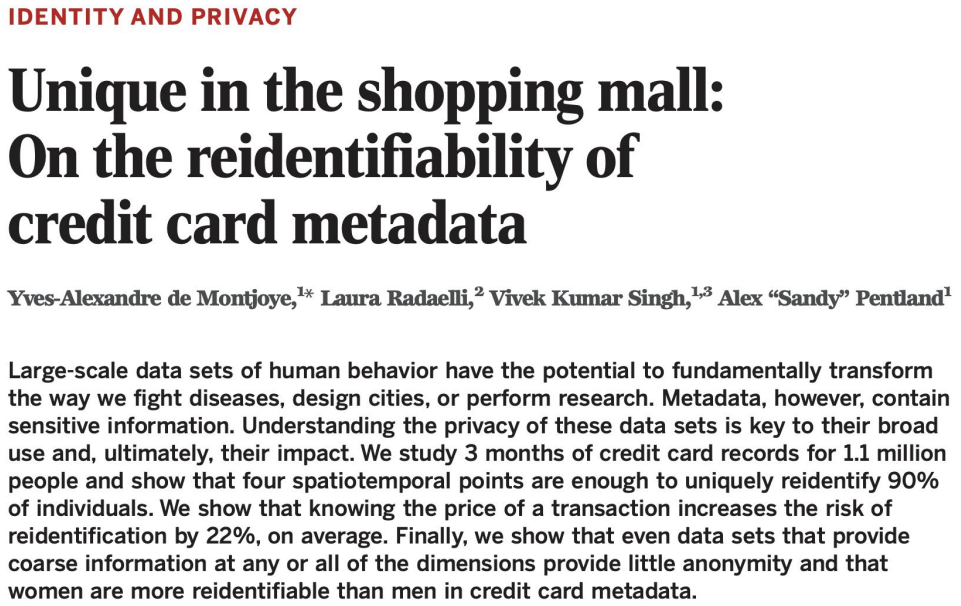

Google Is Absolutely Listening to Your Conversations, and It Confirms Why People Don't Trust Big Tech

In a blog post, the company revealed that audio of Google Assistant conversations is reviewed by humans.

3

Inherent ML Risks

- Linkage Attacks
- Dataset/Feature Reconstruction
- Model Inversion
- Membership Inference
- Attacks Against Models



IDENTITY AND PRIVACY

Unique in the shopping mall: On the reidentifiability of credit card metadata

Yves-Alexandre de Montjoye,[1*] Laura Radaelli,[2] Vivek Kumar Singh,[1,3] Alex "Sandy" Pentland[1]

Large-scale data sets of human behavior have the potential to fundamentally transform the way we fight diseases, design cities, or perform research. Metadata, however, contain sensitive information. Understanding the privacy of these data sets is key to their broad use and, ultimately, their impact. We study 3 months of credit card records for 1.1 million people and show that four spatiotemporal points are enough to uniquely reidentify 90% of individuals. We show that knowing the price of a transaction increases the risk of reidentification by 22%, on average. Finally, we show that even data sets that provide coarse information at any or all of the dimensions provide little anonymity and that women are more reidentifiable than men in credit card metadata.

4

Foster Collaboration

EQUIDEUM HEALTH

Banks

Social Networks

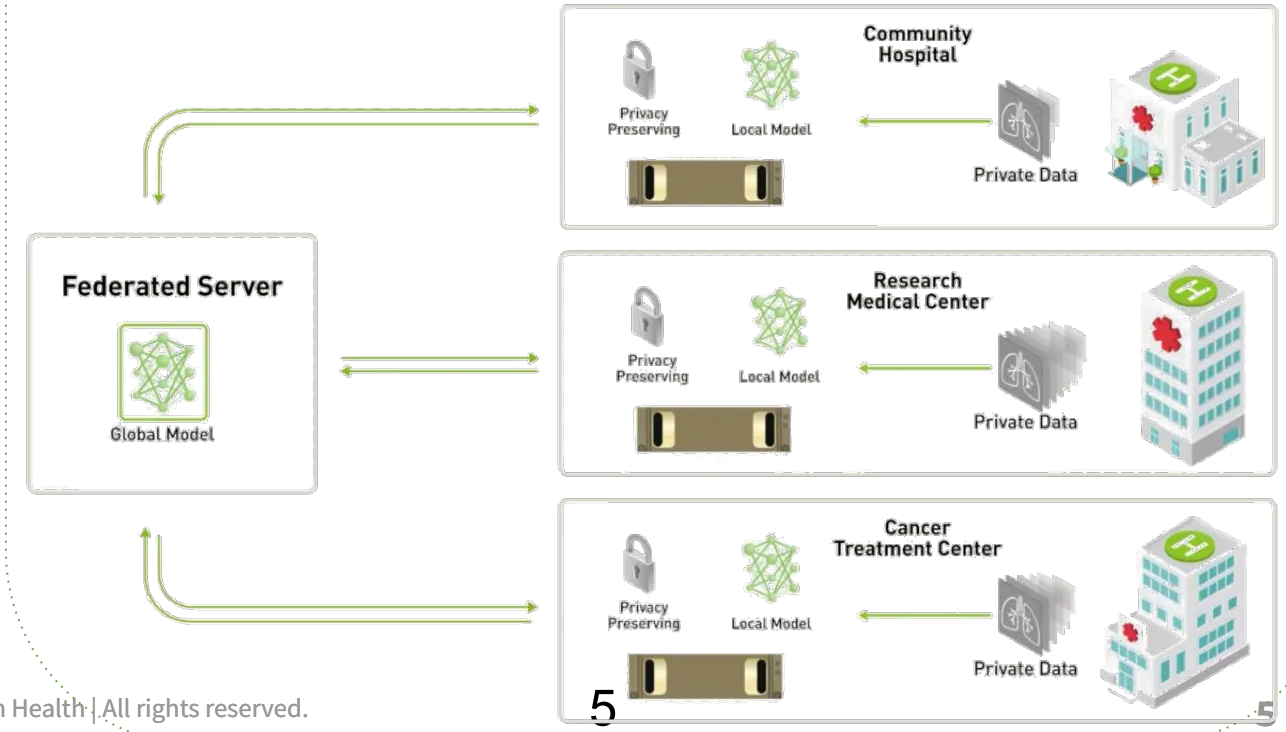
Federated Learning

Smart Consumer Finance

Consumption

Finance

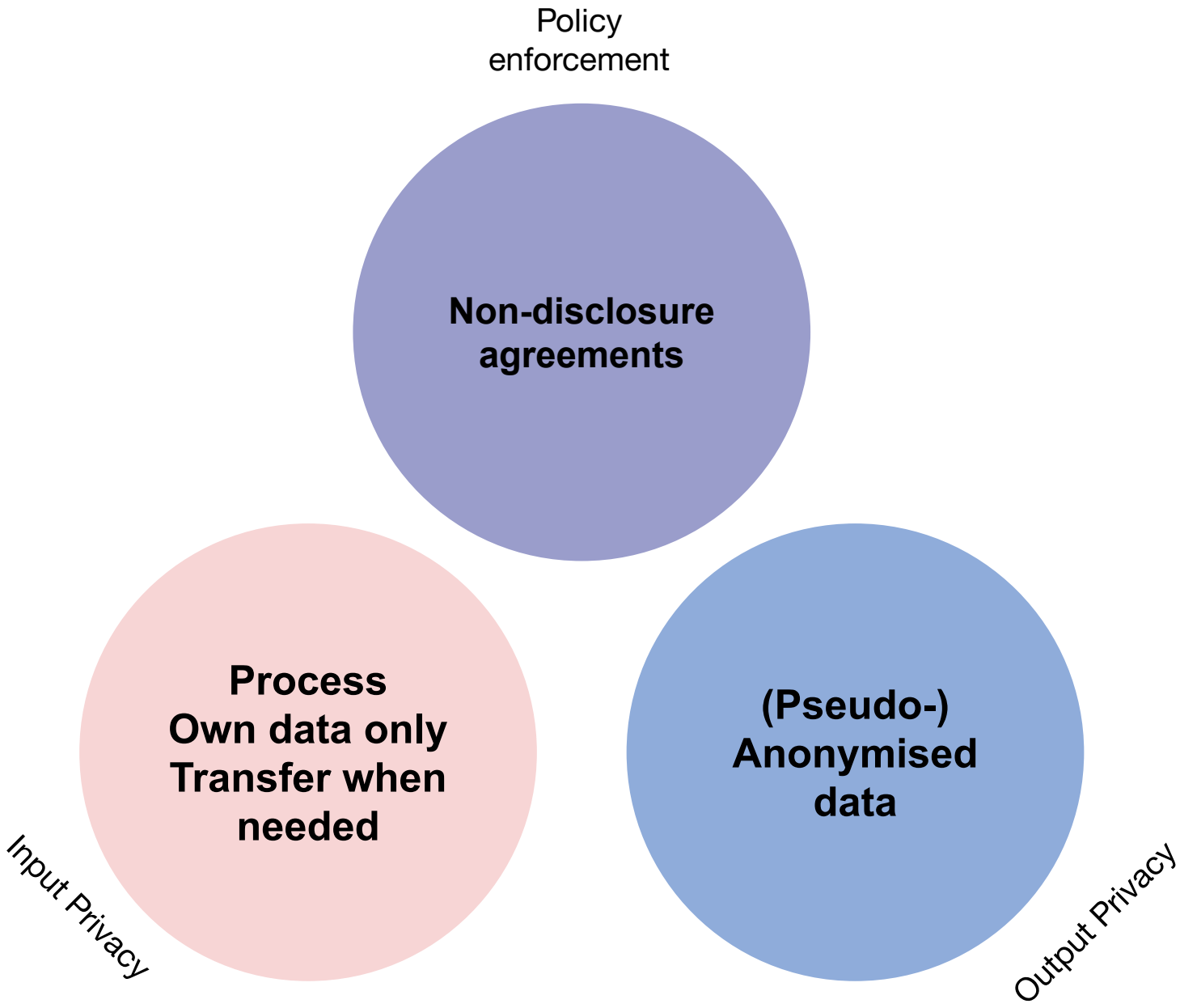
Healthcare



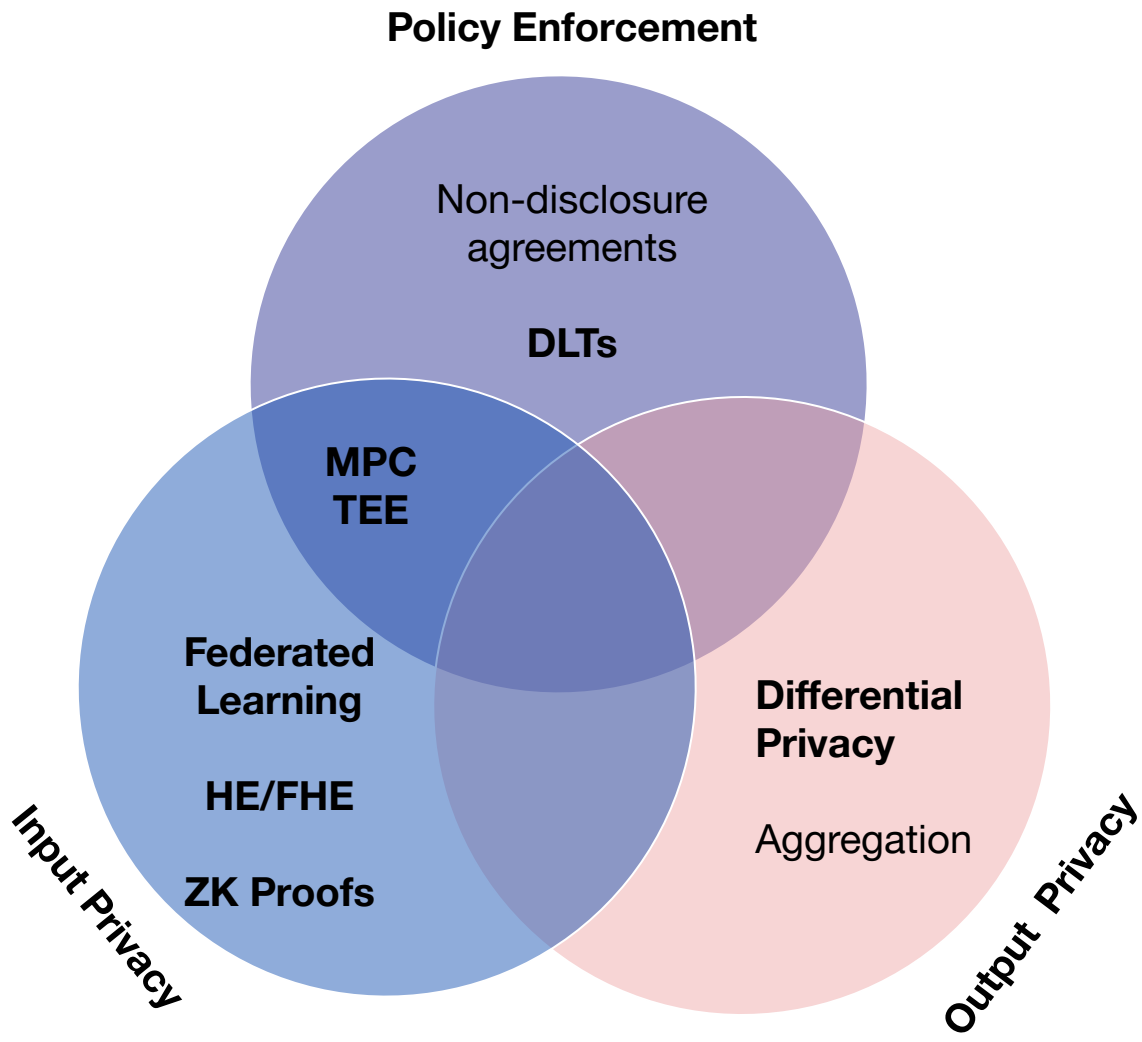
CONSENSYS HEALTH

How:
Privacy Enhancing Technologies

Legacy sensitive data management: **Limited possibilities** and **no synergy**



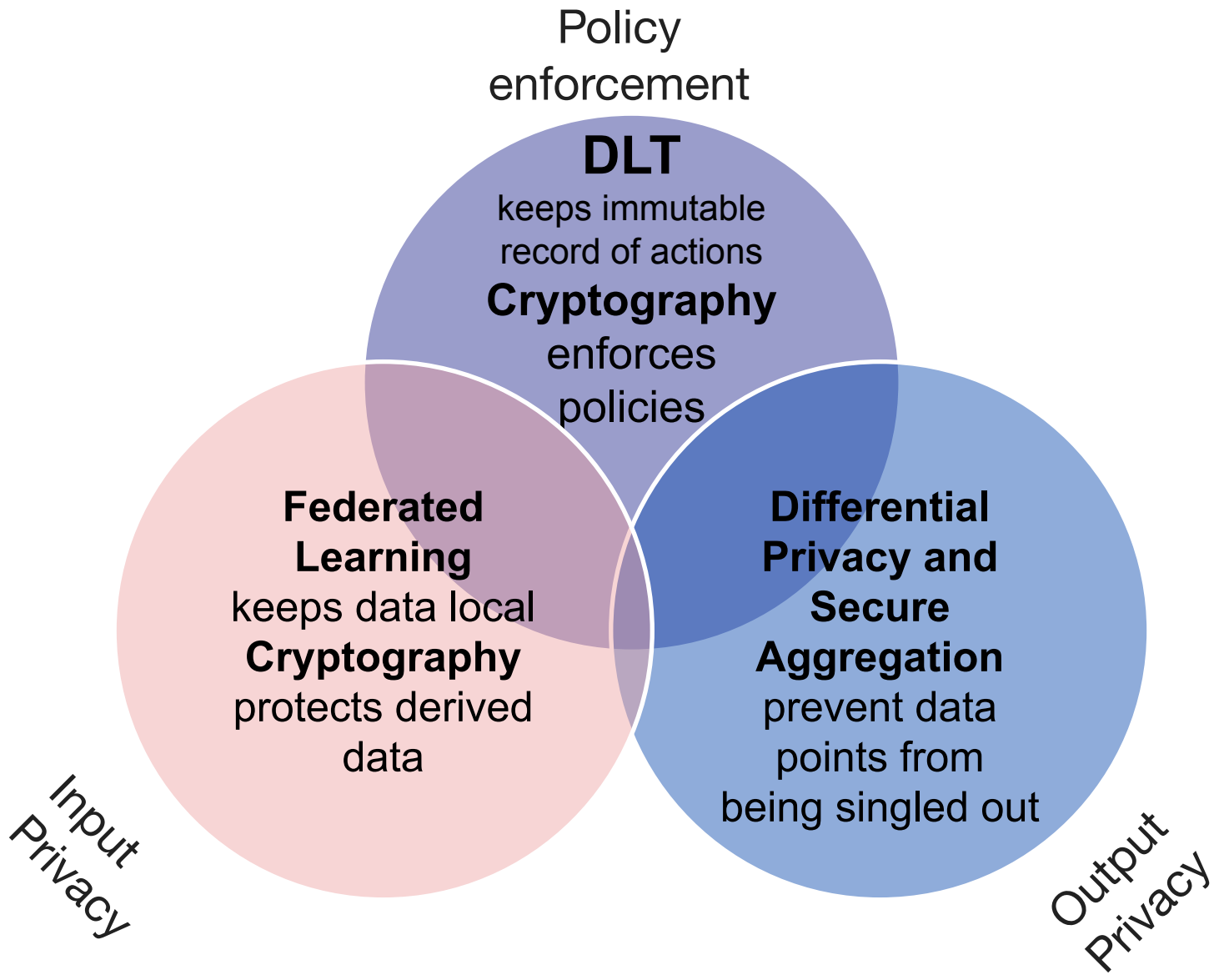
Privacy Enhancing Technologies - 3 Categories



[Credits] UN Handbook on Privacy-Preserving Computation Techniques

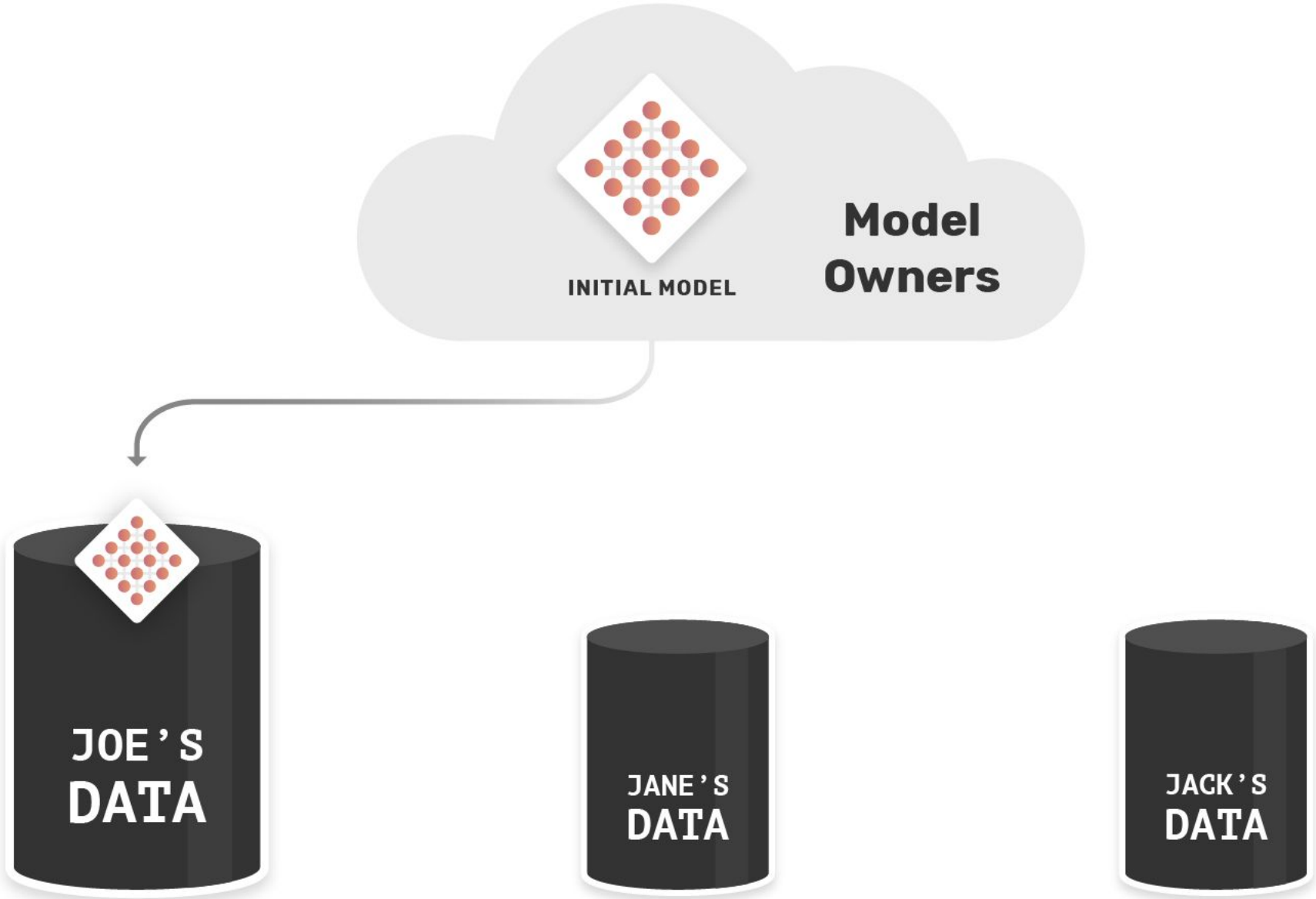
EQUIDEUM HEALTH

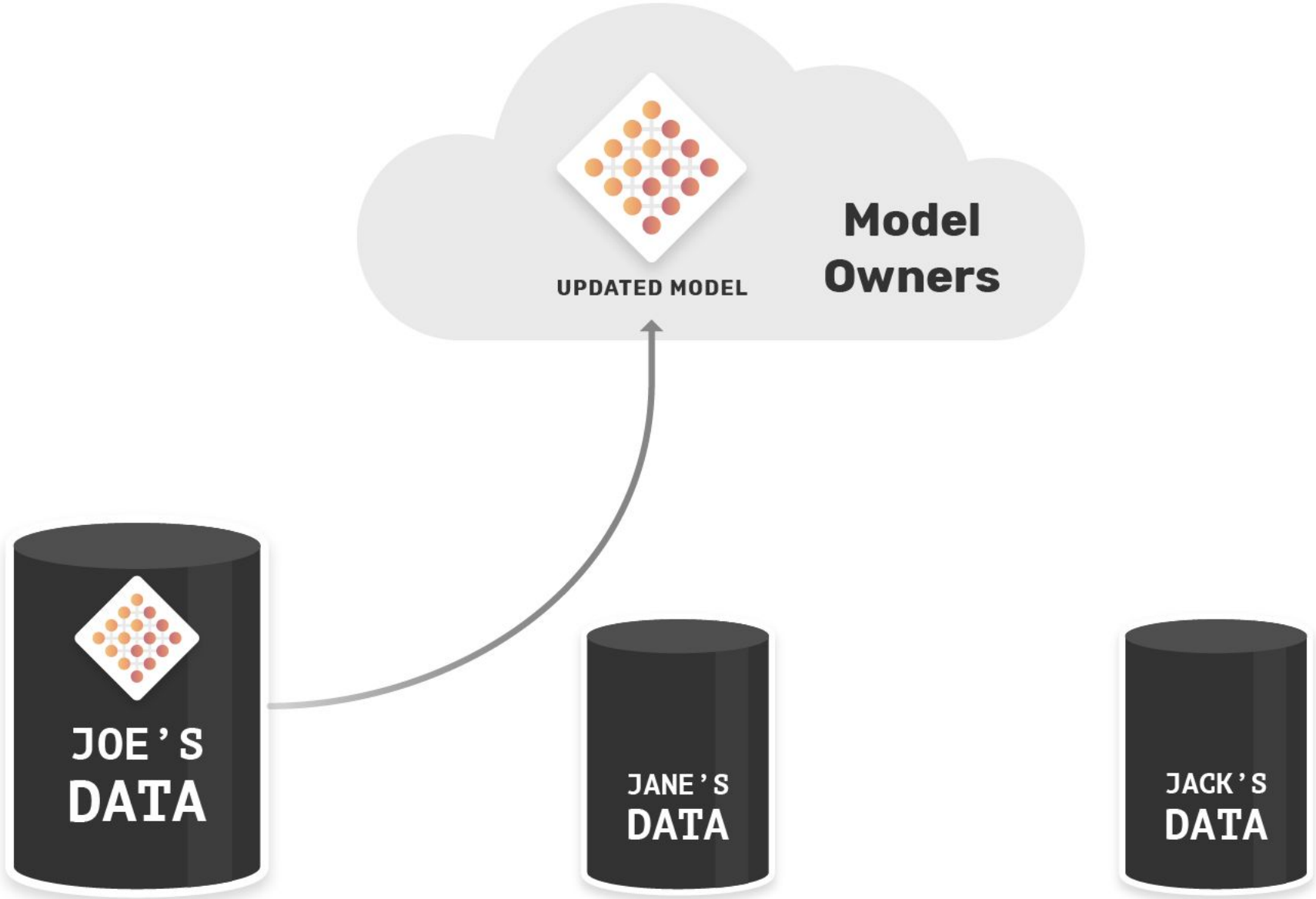
Combination of **Privacy Enhancing Technologies**

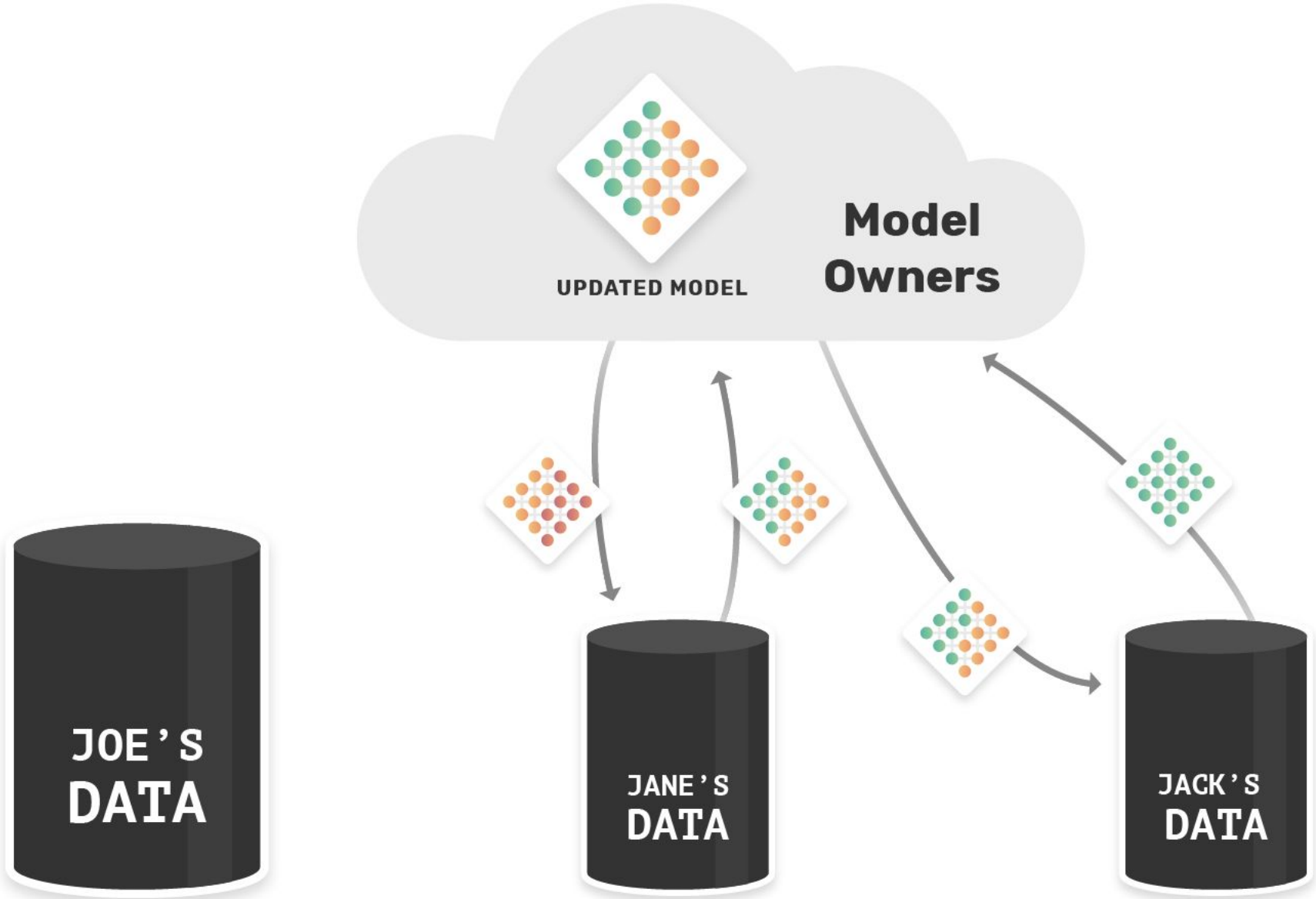


Federated Learning

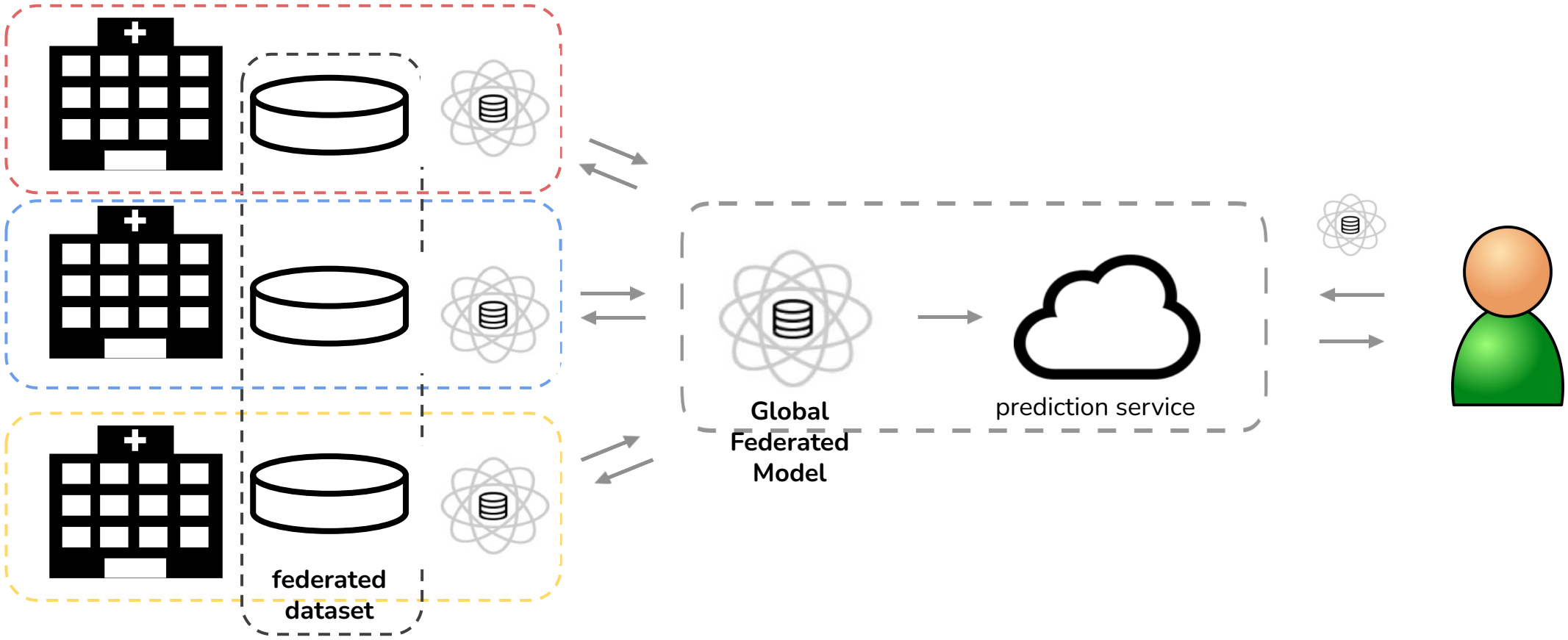
Send the model, not the data.







Collaborative *learning* -> creating better prediction services



Real-world Applications

The Astrix Blog

Privacy Resources Business brave Download B

Using Federated Learning to Improve Brave's On-Device Recommendations While Protecting Your Privacy

Announcements , Community

MELLODDY Consortium Employs Federated Learning and Blockchain to Enhance AI Drug Discovery

MIT Technology Review

Featured Topics Newsletters Events Podcasts

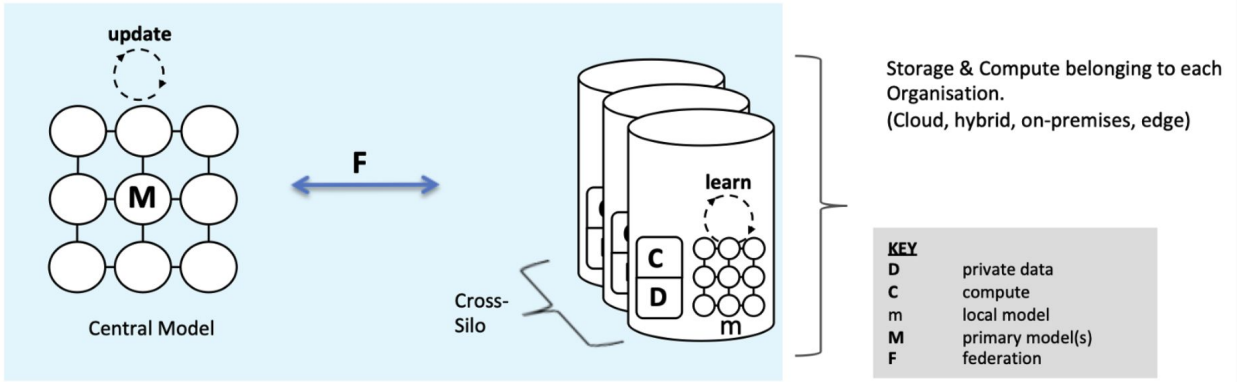
- Google's gBoard (next word, emoji, …
- Also multiple large-scale research initiatives led by Intel, NVIDIA, …

ARTIFICIAL INTELLIGENCE

How Apple personalizes Siri without hoovering up your data

The tech giant is using privacy-preserving machine learning to improve its voice assistant while keeping your data on your phone.

Cross-silo vs Cross-device



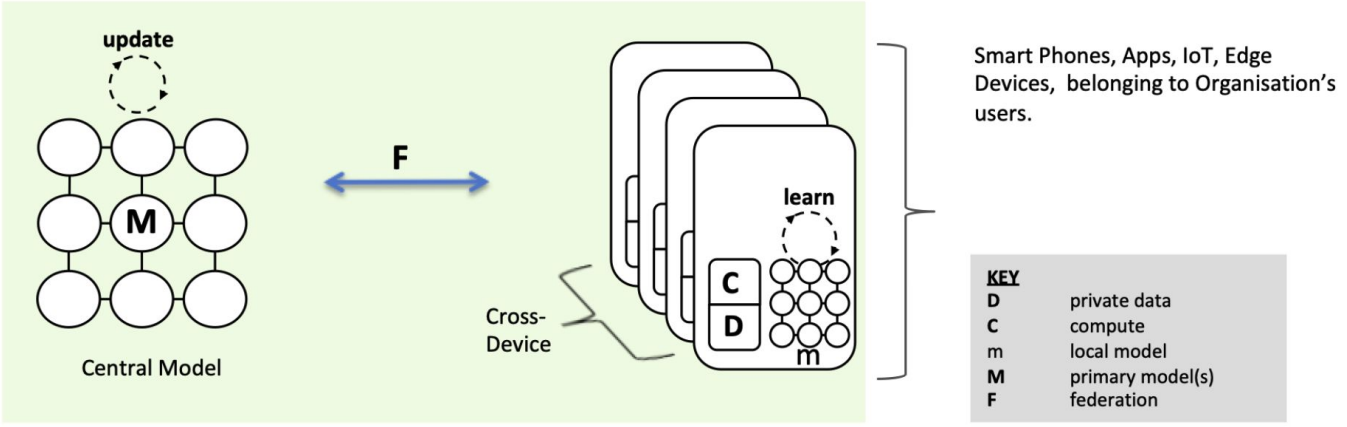
Multiple Organisations, Model Centric & Cross-Silo Federated Learning

Cross-Silo FL

- More reliable compute and network (servers)
- Larger data
- Less clients

Cross-Device FL

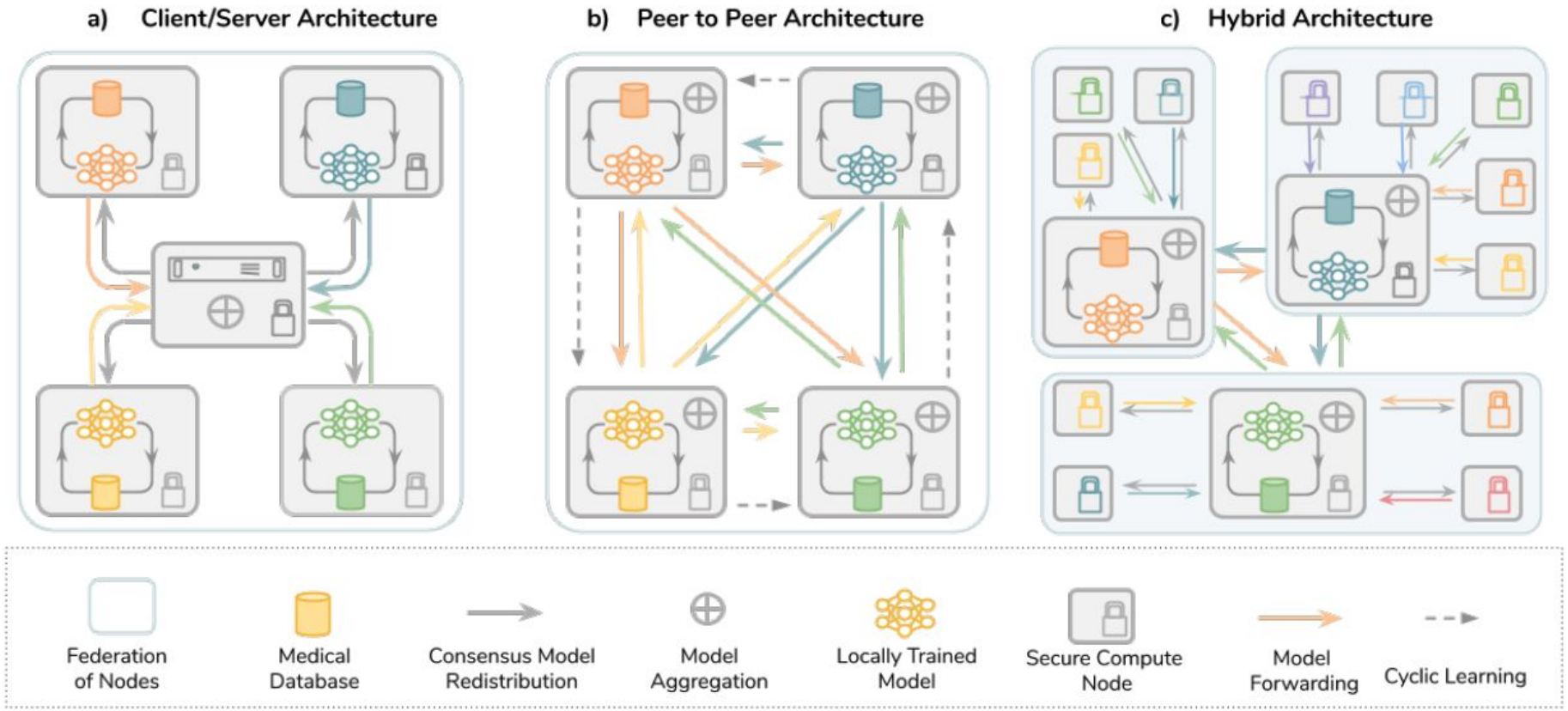
- Network and devices not reliable - Usually mobile devices (smartphones / IOT)
- Smaller data
- Large number of clients



Single Organisation, Model-Centric & Cross-Device Federated Learning

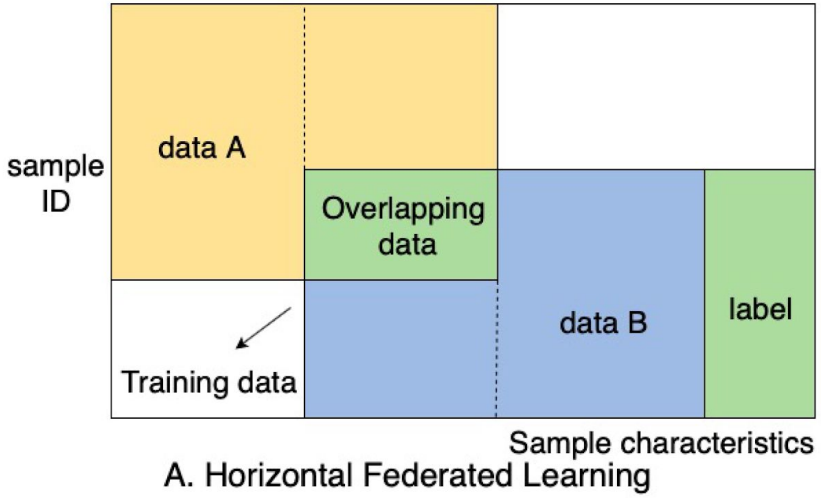
https://blog.openmined.org/federated-learning-types/

FL architectures



Types of FL

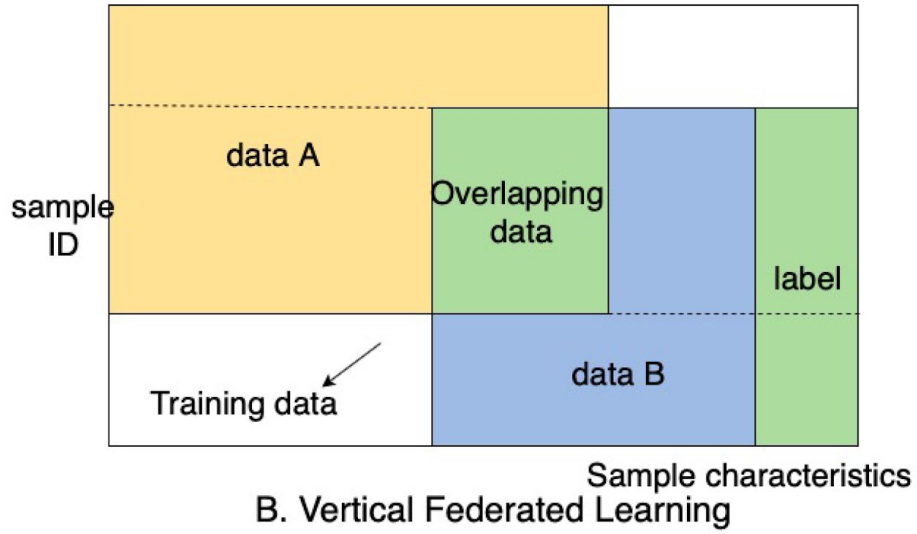
Horizontal FL



A. Horizontal Federated Learning

Shared **feature** space

Vertical FL



B. Vertical Federated Learning

Shared **sample** space

Software frameworks

Simulation frameworks

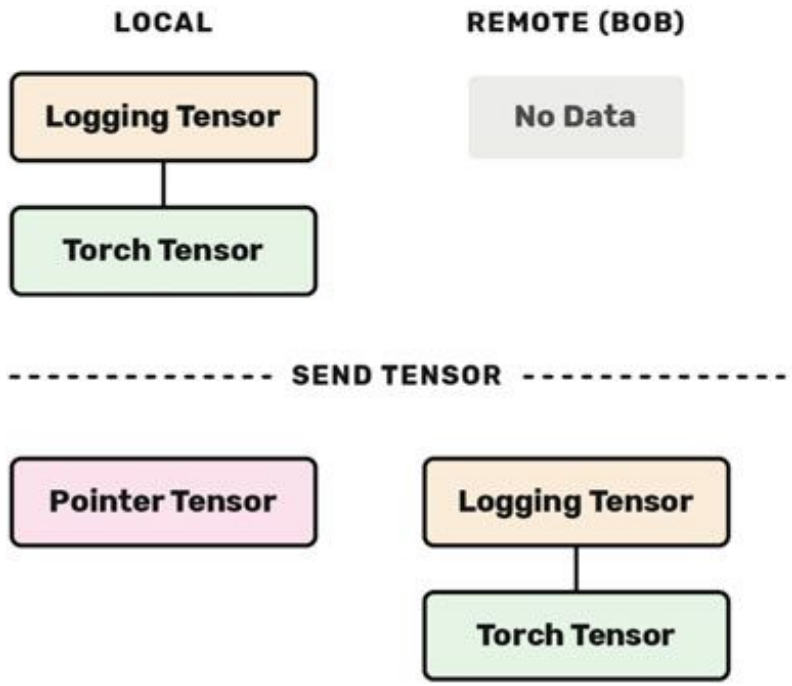
- TF Federated
- Microsoft FLUTE
- Meta's FLSim

Actual distributed Computing

- Flower
- Intel OpenFL
- Substra
- FATE
- FedML

Remote data science

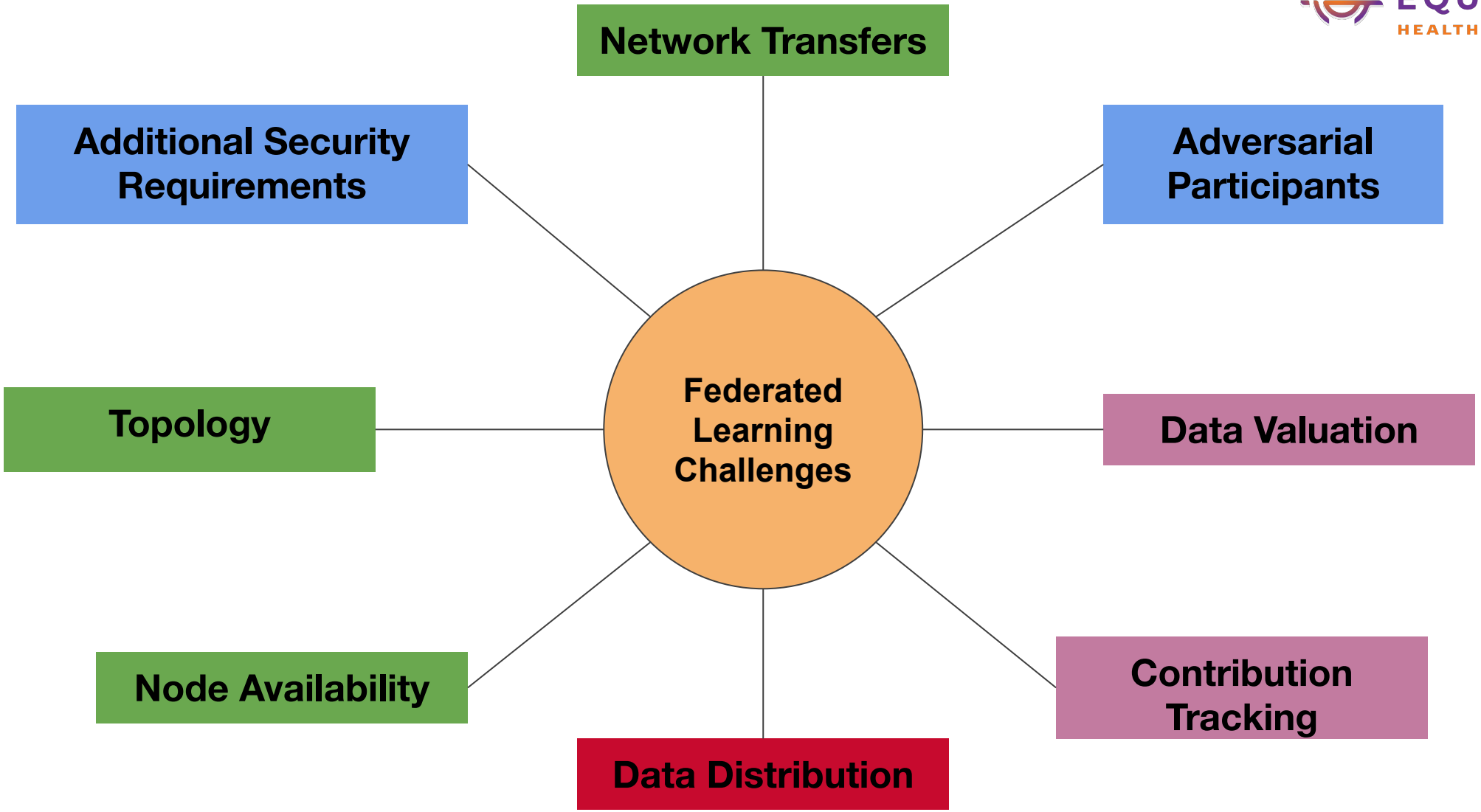
- PySyft



Federated Learning

Technical Challenges and Research problems

EQUIDEUM HEALTH



ML SysOps DLT / Tokenisation Security & Privacy

FL Challenges

- **Non-IID** data - Training data for a given client is typically site specific, hence the site's local dataset will not be representative of the distribution of training samples.
- **Unbalanced data** - Sites may have a lot or little training data, leading to varying amounts of local training data across different sites.
- **Massively distributed** data - There may be extreme scenarios where each site only has very few training samples (in the limiting case one example)
- **Communication costs** - Communication between clients and servers occurs communication overheads. Depends on the number of clients and the frequency of updates from/to server.

Where do these challenges come from?

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)

Cross-device

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**

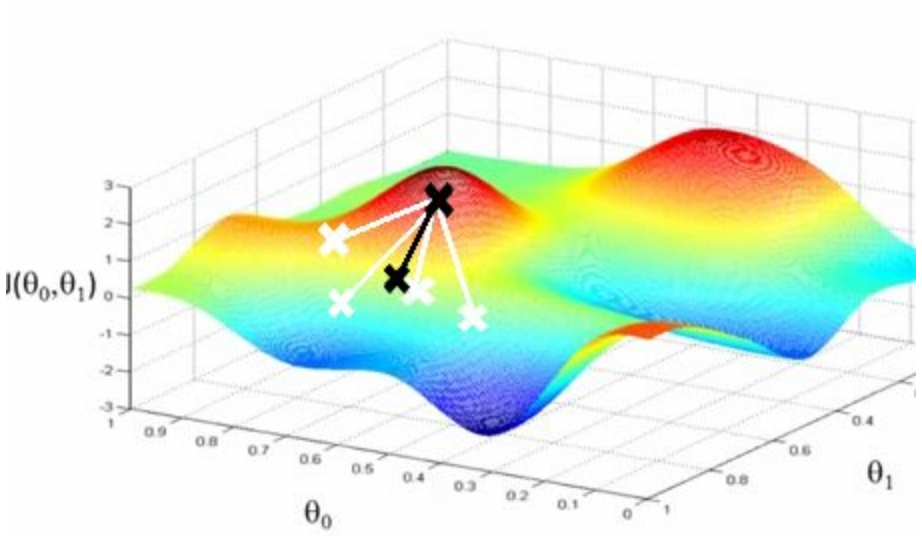
$w \leftarrow w - \eta \nabla \ell(w; b)$

Local update
Local objective…

return w to server

https://arxiv.org/pdf/1806.00582.pdf

IID vs non-IID data



IID vs non-IID data

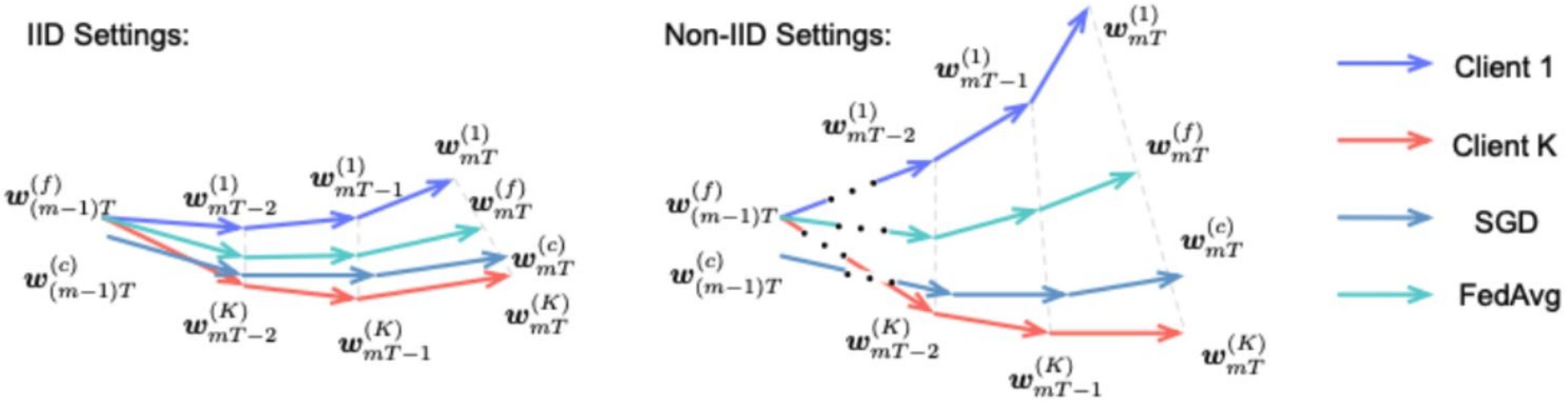
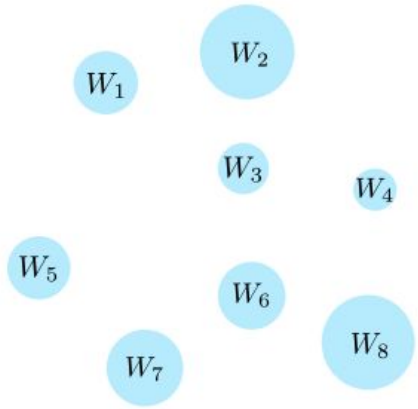


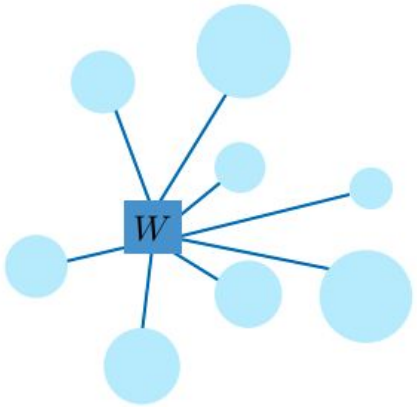
Figure 3: Illustration of the weight divergence for federated learning with IID and non-IID data.

https://arxiv.org/pdf/1806.00582.pdf

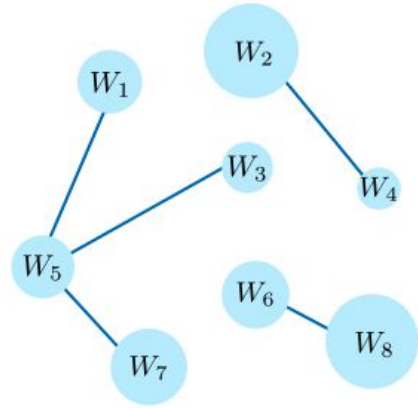
Personalisation



(a) Learn personalized models for each device; do not learn from peers.



(b) Learn a global model; learn from peers.

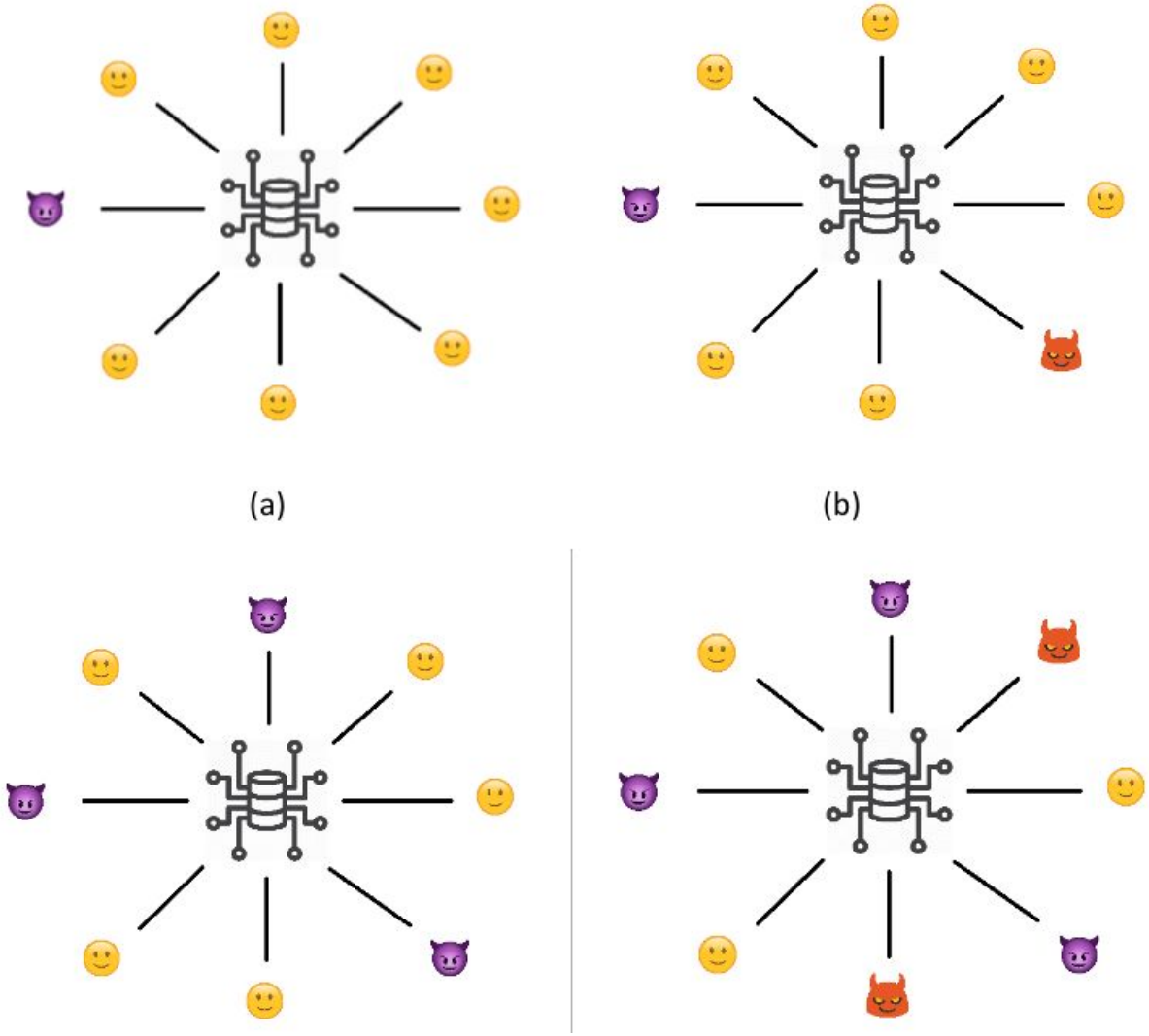


(c) Learn personalized models for each device; learn from peers.

Figure 5: Different modeling approaches in federated networks. Depending on properties of the data, network, and application of interest, one may choose to (a) learn separate models for each device, (b) fit a single global model to all devices, or (c) learn related but distinct models in the network.

https://arxiv.org/pdf/1908.07873.pdf

Robust Aggregation - Adversaries



Robust Aggregation - Mitigations

Goals:

Adversarial Defense, Non-IID mitigation, Privacy, Personalization

Methods:

- **Norm Clipping**: Robustness to outliers
- **Zeroing**: Robustness to corrupted data
- **Weighted Averaging**: Multiple domain adaptation
- **Differential Privacy:** Minimize information Leakage
- **Learning-based:** detect adversary contributions
- **Advanced cryptography:** Clients can prove facts about their contributions

Weights vs gradients

- Send **weights** to avoid sending multiple **gradients** and save communication costs
- Weights carry less information => easier to defend against adversaries

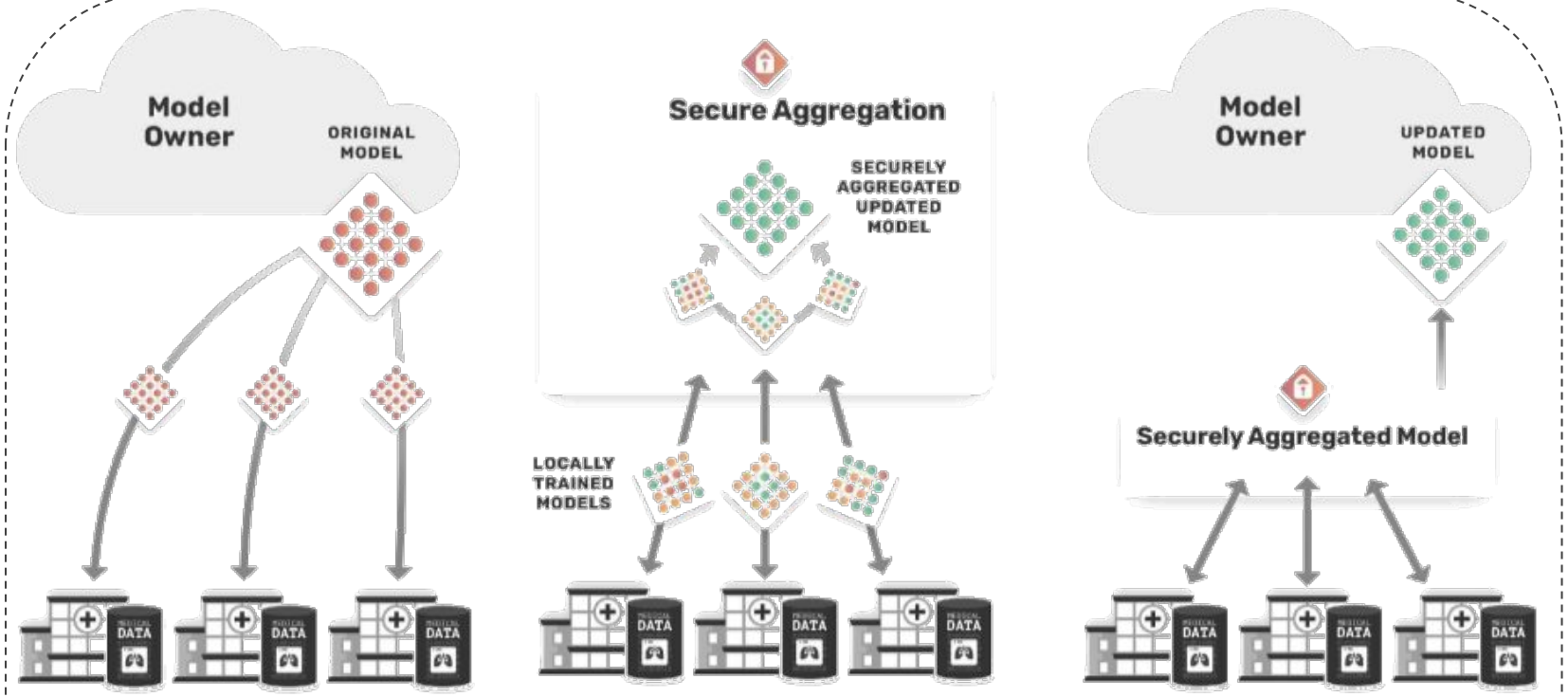
Aggregation methods

Goals:

- Robustness
- Privacy
- Compression
- Security
- Personalization

Secure Aggregation

***A** - One round of Federated Learning*

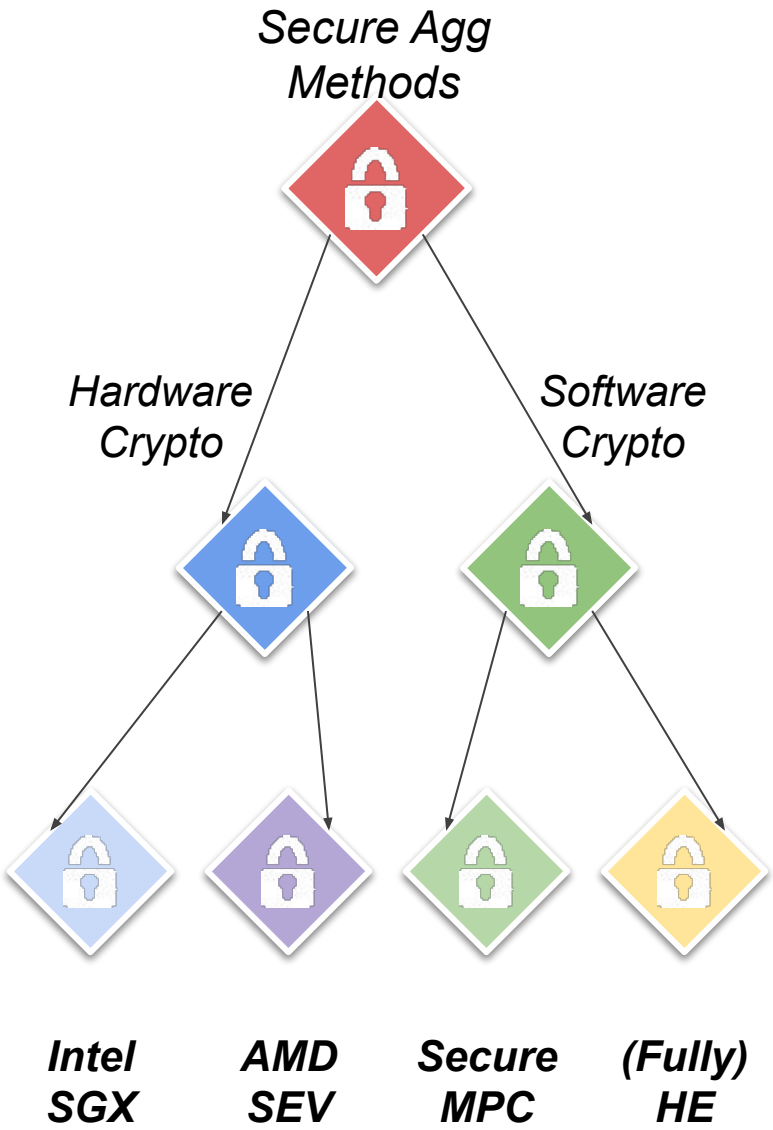


Model sent to each site for **training on local raw data** ①

Local updates (not raw data) encrypted and sent back for aggregation ②

Securely aggregated model sent back to local sites and model owner ③

***B** - Restart from 1 for full Federated Learning Protocol*



Research challenges

- **Representative datasets (IID vs non-IID)**
- Evaluating contributions
- Guaranteeing **privacy**
  - How much privacy do we need?
  - Can we relax some assumptions depending on the context?
- Fair valuation of data/contributions

Compliance with regulations?

Inherent ML Risks

Attacks against models / data

- Linkage Attacks
- Dataset/Feature Reconstruction
- Model Inversion
- Membership Inference

IDENTITY AND PRIVACY

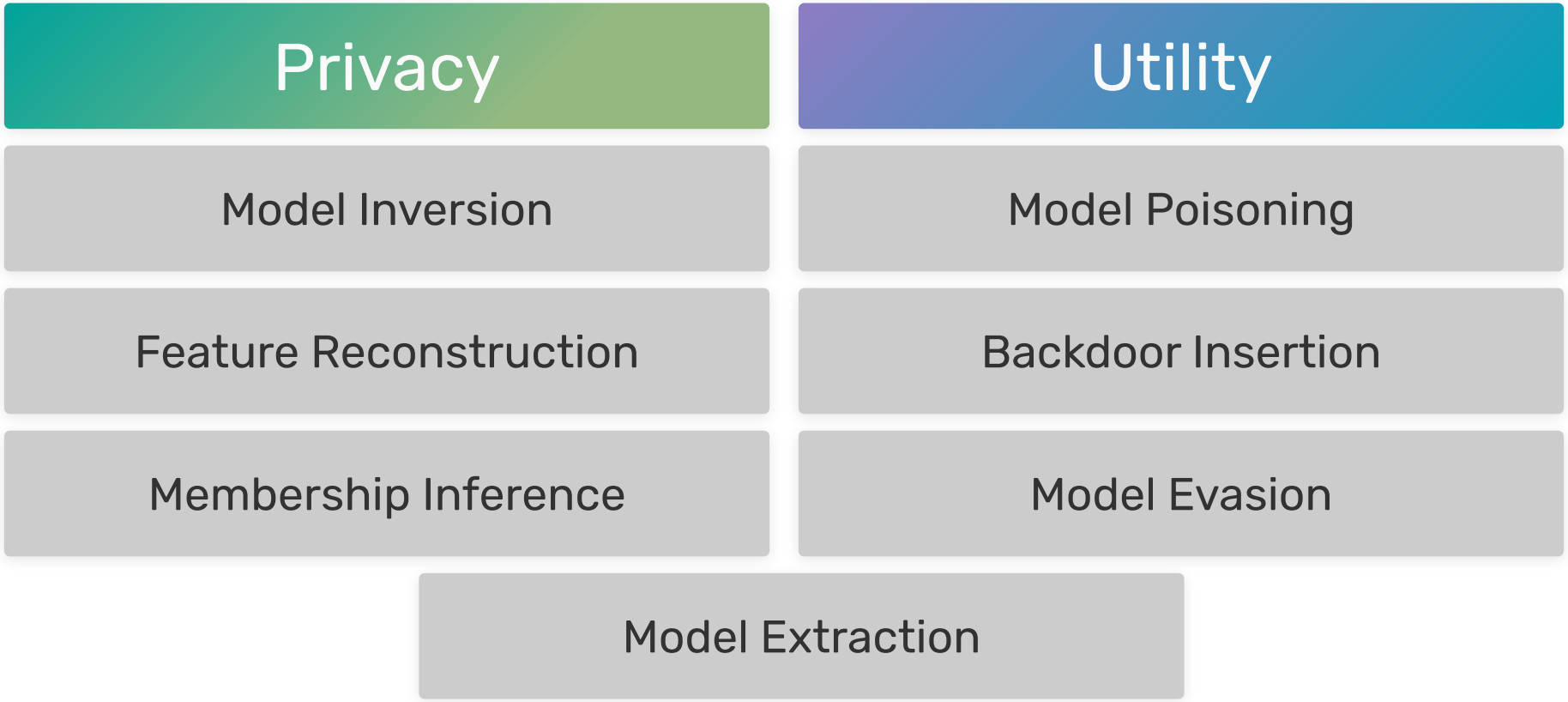
Unique in the shopping mall: On the reidentifiability of credit card metadata

Yves-Alexandre de Montjoye,[1*] Laura Radaelli,[2] Vivek Kumar Singh,[1,3] Alex "Sandy" Pentland[1]

Large-scale data sets of human behavior have the potential to fundamentally transform the way we fight diseases, design cities, or perform research. Metadata, however, contain sensitive information. Understanding the privacy of these data sets is key to their broad use and, ultimately, their impact. We study 3 months of credit card records for 1.1 million people and show that four spatiotemporal points are enough to uniquely reidentify 90% of individuals. We show that knowing the price of a transaction increases the risk of reidentification by 22%, on average. Finally, we show that even data sets that provide coarse information at any or all of the dimensions provide little anonymity and that women are more reidentifiable than men in credit card metadata.

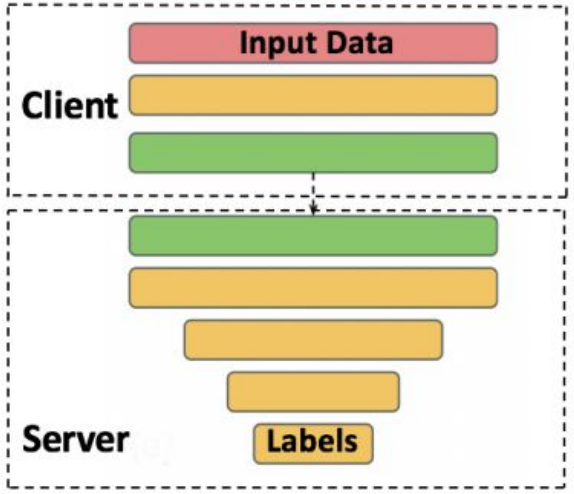
- All this **while respecting CCPA, GDPR, ...**

Attacks against collaborative learning systems

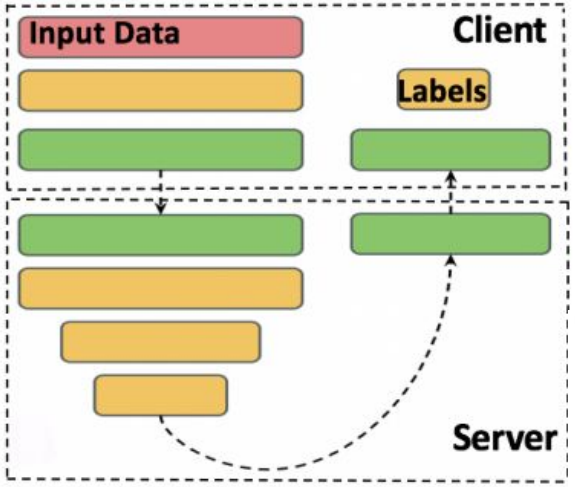


Usynin, D., *et al.* Adversarial interference and its mitigations in privacy-preserving collaborative machine learning. *Nat Mach Intell* https://doi.org/10.1038/s42256-021-00390-3

Other collaborative learning methods
SplitNN



(a) Vanilla split learning



(b) U-shaped split learning

QUANTITATIVE RESULTS

Method	100 Clients	500 Clients
Large Scale SGD	29.4 TFlops	5.89 TFlops
Federated Learning	29.4 TFlops	5.89 TFlops
Our Method (SplitNN)	0.1548 TFlops	0.03 TFlops

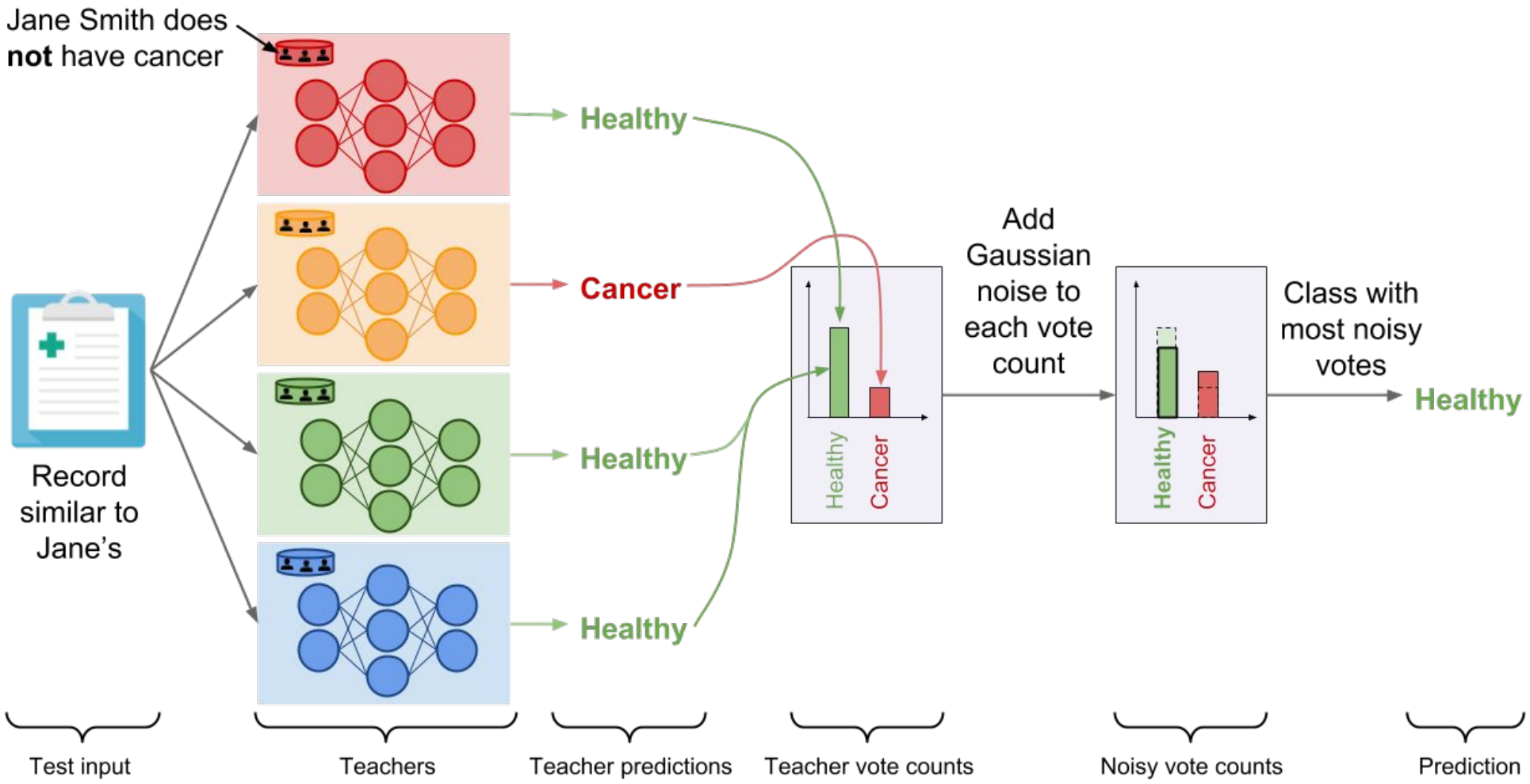
Table 1. Computation resources consumed per client when training CIFAR 10 over VGG (in teraflops)

Method	100 Clients	500 Clients
Large Scale SGD	13 GB	14 GB
Federated Learning	3 GB	2.4 GB
Our Method (SplitNN)	6 GB	1.2 GB

Table 2. Communication Bandwidth consumed per client when training CIFAR 100 and Resnet 50 (in gigabytes)

https://www.media.mit.edu/projects/distributed-learning-and-collaborative-learning-1/overview/

Other collaborative learning methods
PATE



Differential Privacy

Make statistical deductions about data without disclosing individual information

Prevent information leakage

- Data anonymisation is **not an option**
  - Anonymity is not enough: some data can identify uniquely the people
  - > Netflix dataset, 2006: (ratings, dates) is a good identifier
- Differential Privacy (DP)
  - Idea: no single element in the dataset can significantly influence the output.
  - Or, we can't use the frequency specific to one output to distinguish datasets.

Reconstructed Original

Fredrikson et al., 2015

Usynin, D., *et al.* Zen and the art of model adaptation: Low-utility-cost attack mitigations in collaborative machine learning. *Proceedings on Privacy Enhancing Technologies, 2021* (in press)

Differential Privacy

EQUIDEUM HEALTH

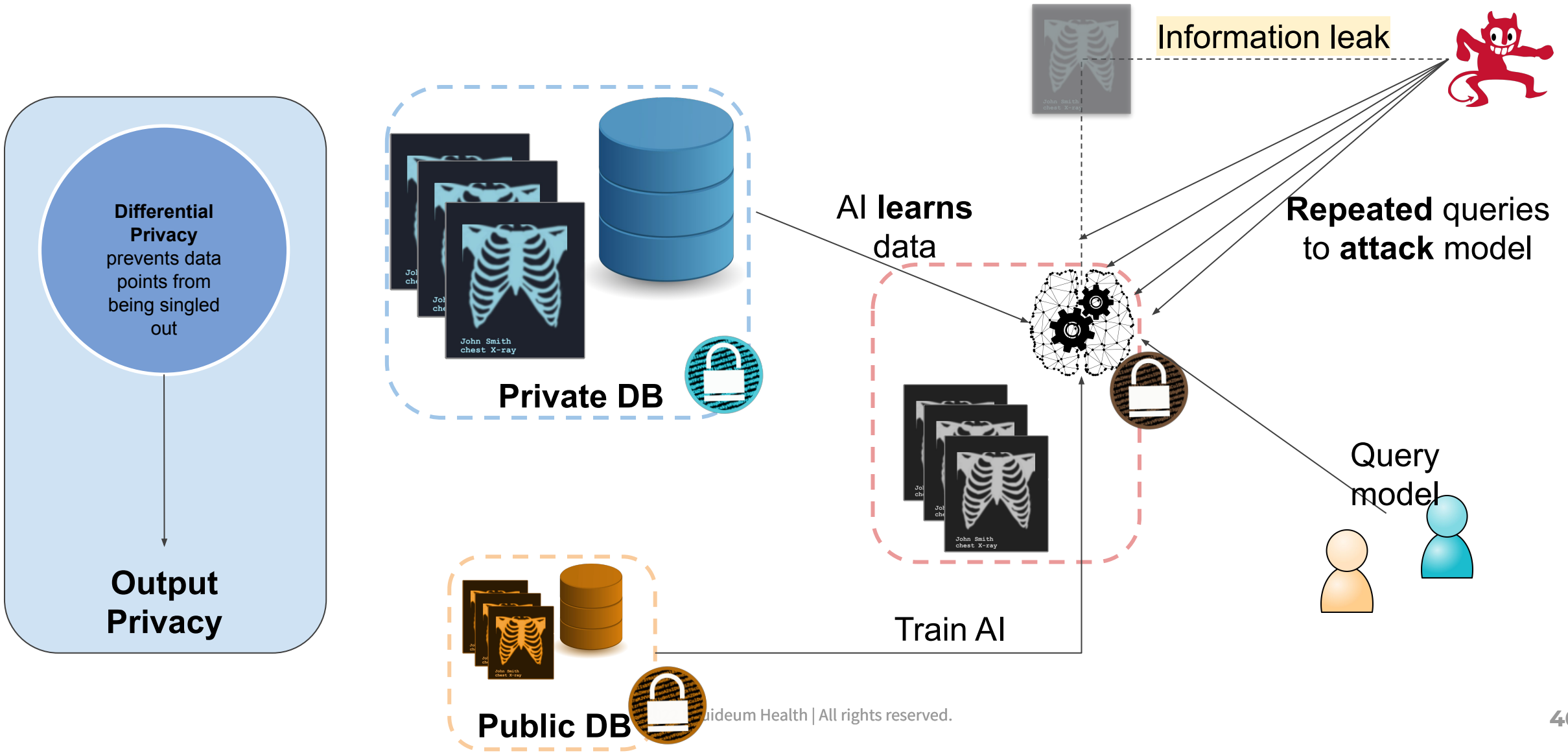
- Property of an **algorithm**
- **Formalise** the concept of privacy and make it quantifiable (*privacy budget*)
- Add controlled amount of **statistical noise**
- **Perfect privacy**: the output of our query is the same between this database and any identical database with one row removed or replaced (*plausible deniability*)

An algorithm $M : X^n \to Y$ is DP if, for all neighbouring datasets X and X' and all $T \subseteq Y$:

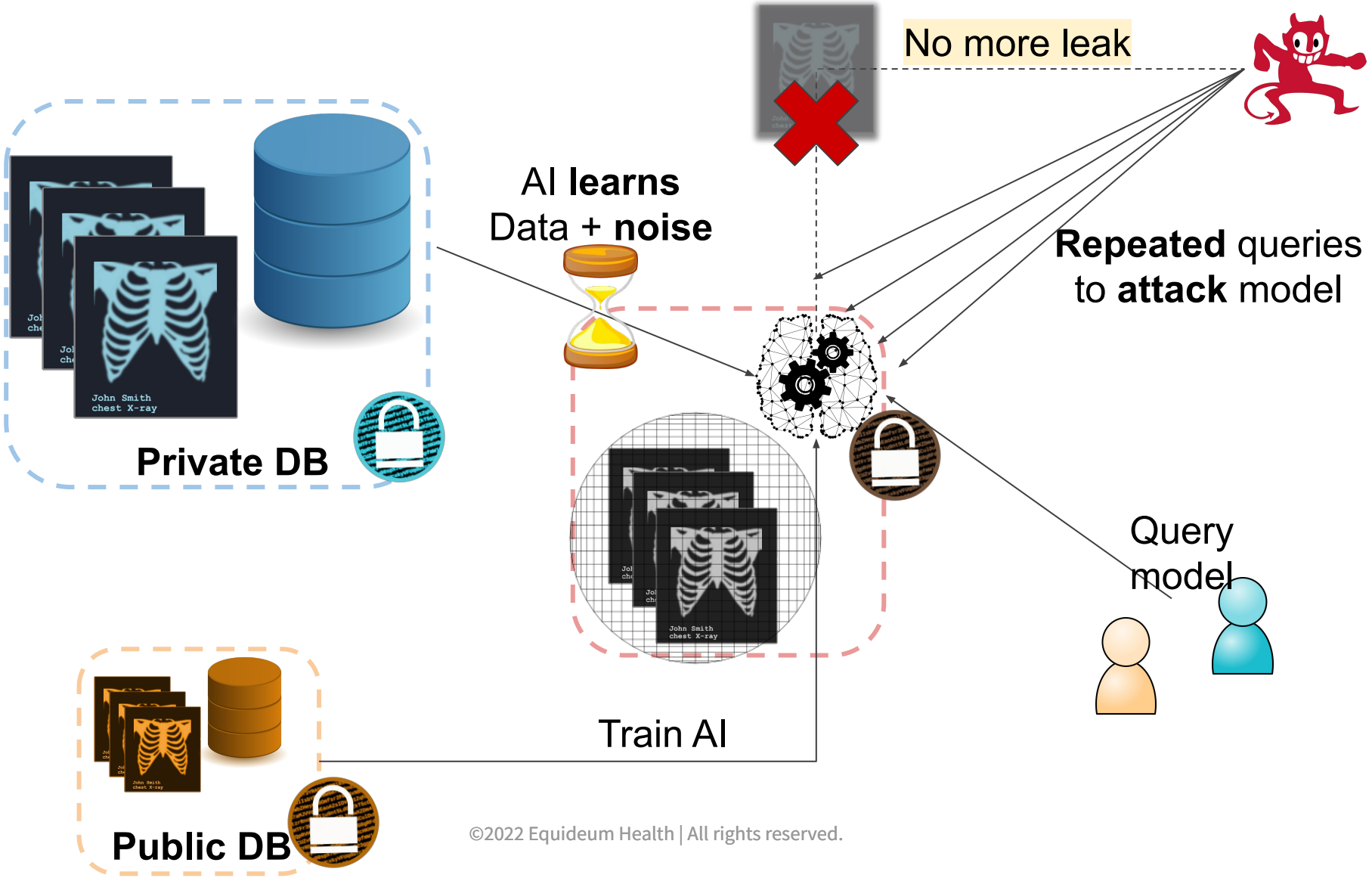
$$P[M(X) \in T] \leq e^{\epsilon} P[M(X') \in T]$$

Dwork, et al. 2006

How can information be leaked?



Differential Privacy



Differential Privacy Meets Deep Learning

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.

Differential Privacy - Challenges

- Translate between privacy budget and business objectives
- Methodological implementations and mode of application (local, global…)
- Privacy-utility trade-off
- Gives an upper bound, how close?
- Compatibilities with existing ML/Deep Learning (BatchNorm, …)

Differential Privacy

- Academic references
  - **Abadi et al.,** *Deep learning with differential privacy*, 2016
  - **Papernot et al.**, *Semi-supervised knowledge transfer for deep learning from private training data*, 2016
  - **Dwork et al.**, *The algorithmic foundations of differential privacy*, 2014
- Software implementations
  - Facebook's Opacus https://github.com/pytorch/opacus
  - Google's TF Privacy https://github.com/tensorflow/privacy

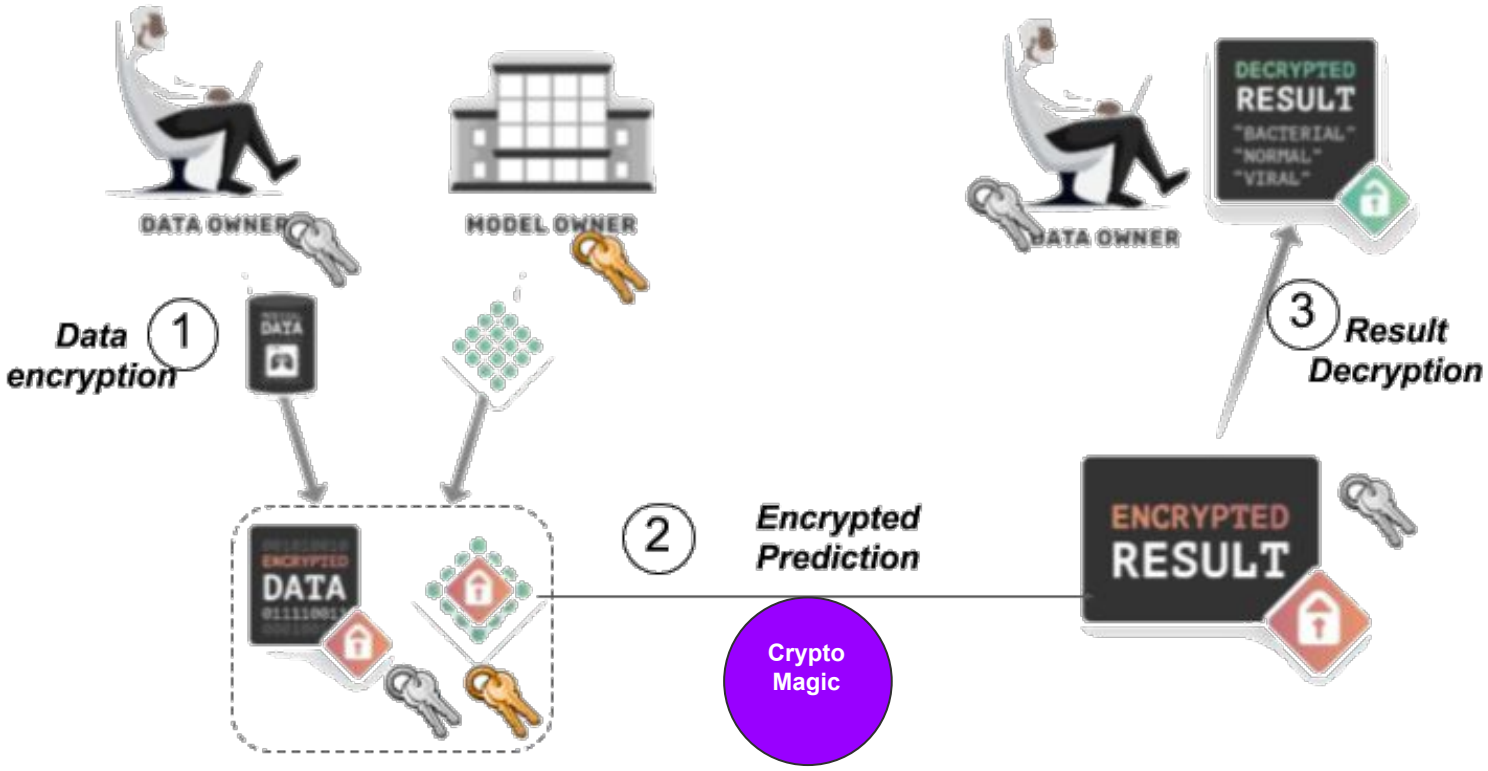
Secure Computation

Allows computations to be performed on data while it remains encrypted/hidden

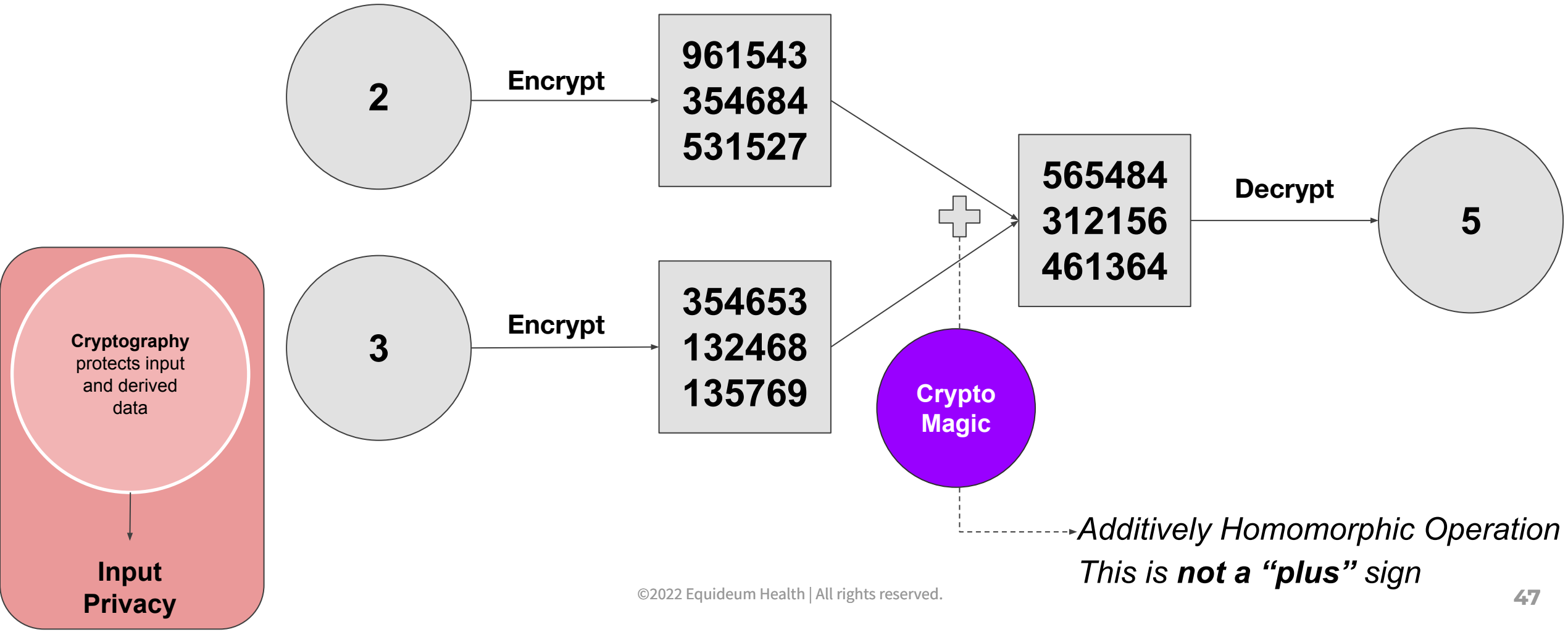
Confidential Compute

Private/Secure Compute (*also Confidential*)

- Individual or group want to **process sensitive data** using a 3rd party service
- Service provider concerned about IP *(think expensive ML model)*



Homomorphic Encryption

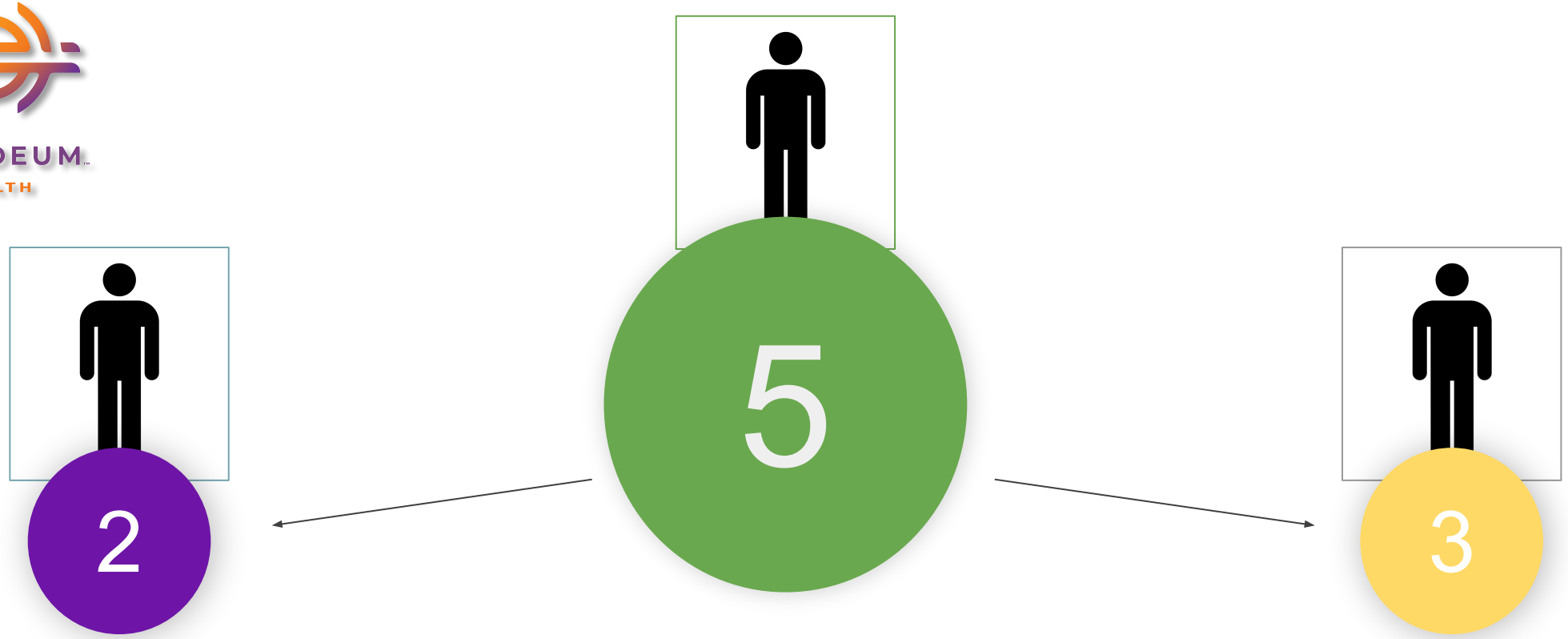


Homomorphic Encryption

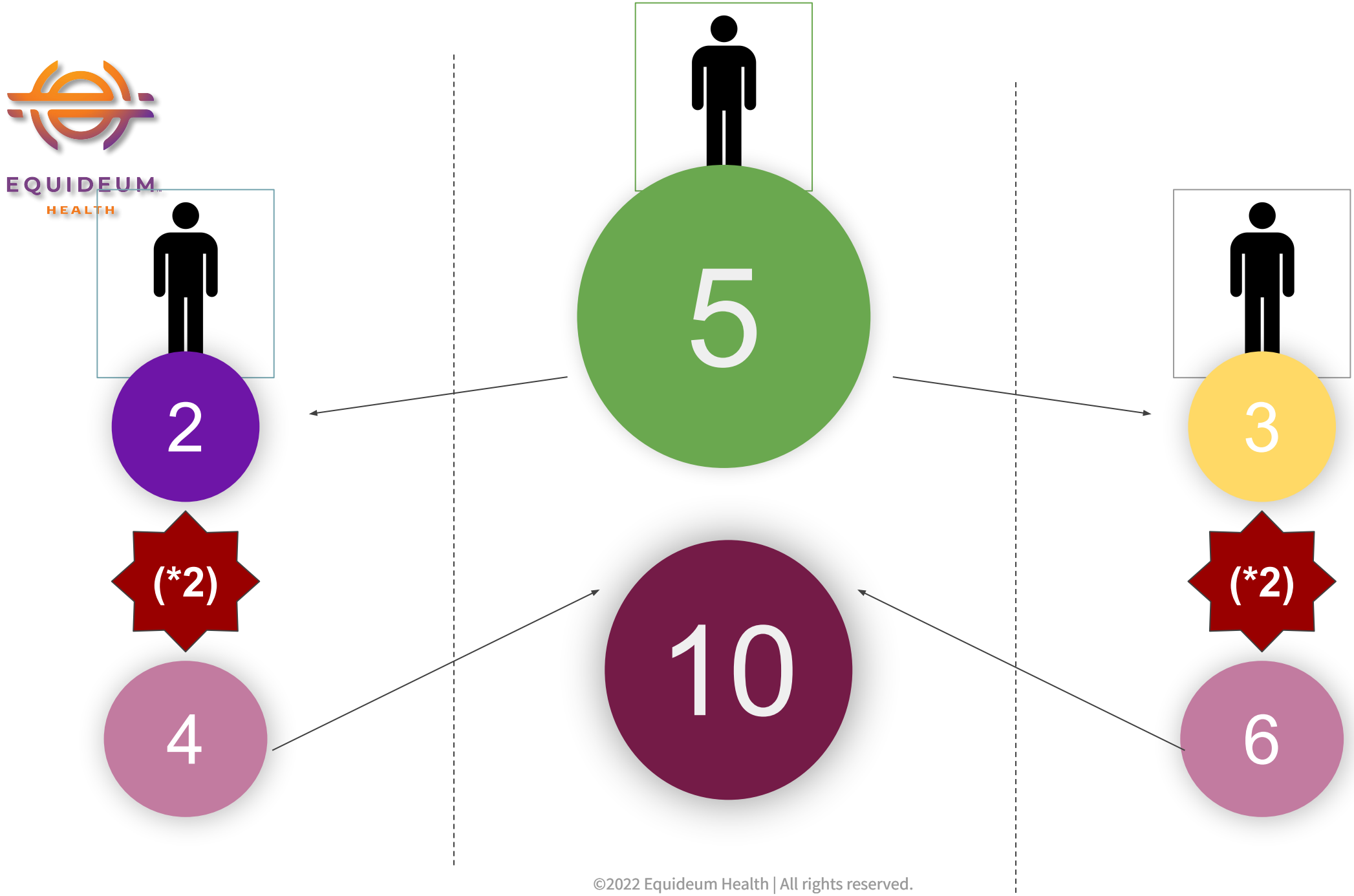
- Extension of symmetric or public-key crypto to computation without key access
- Schemes:
  - **Partially**: Single-gate circuits (addition, multiplication…)
  - **Somewhat**: >1 gate, only a subset of possible circuits
  - **Fully**: arbitrary circuits of unbounded depth
- **Multiplicative depth** is the main practical limitation
- Some quantum-safe algorithms ☺️

Frameworks: HElib, SEAL, Concrete

Secure Multi-Party Computation



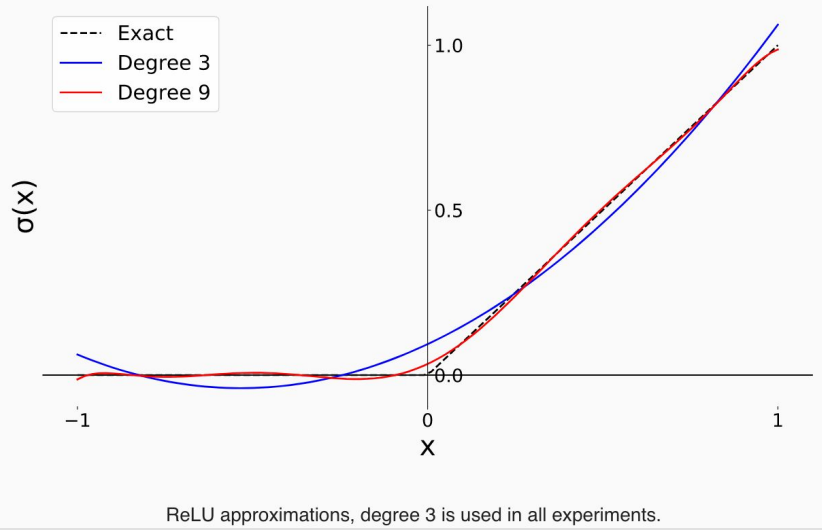
- **Confidentiality:** neither knows the real value
- **Shared Governance**: The number can only be disclosed if everyone agrees



Secure Multi-Party Computation

Operations needed for Deep Learning algorithms

- **Addition & Multiplication**
- **Convolution and Linear layers** → *Only multiplications and additions needed*
- **Pooling** → *Average pooling*
- **Activation: Relu & Sigmoid** → *Polynomial approximations*



ReLU approximations, degree 3 is used in all experiments.

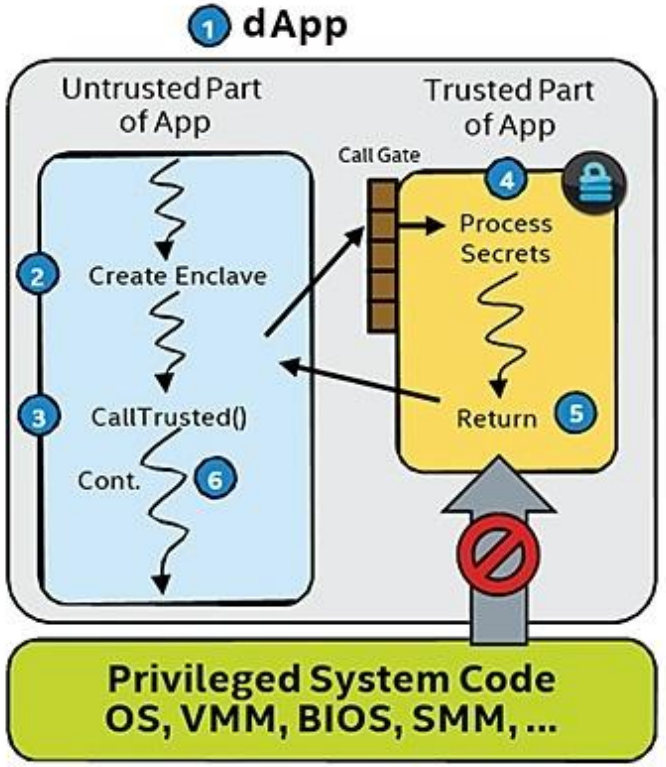
Polynomial approximations of ReLU

[Credits] *https://blog.openmined.org/training-cnns-using-spdz/*

Trusted Execution Environments

- Set of **CPU instructions** to create enclaves in RAM, that **no one can access** - except code from the enclave itself

- Ensures **total confidentiality of data during computation** - decryption happens only inside the enclave

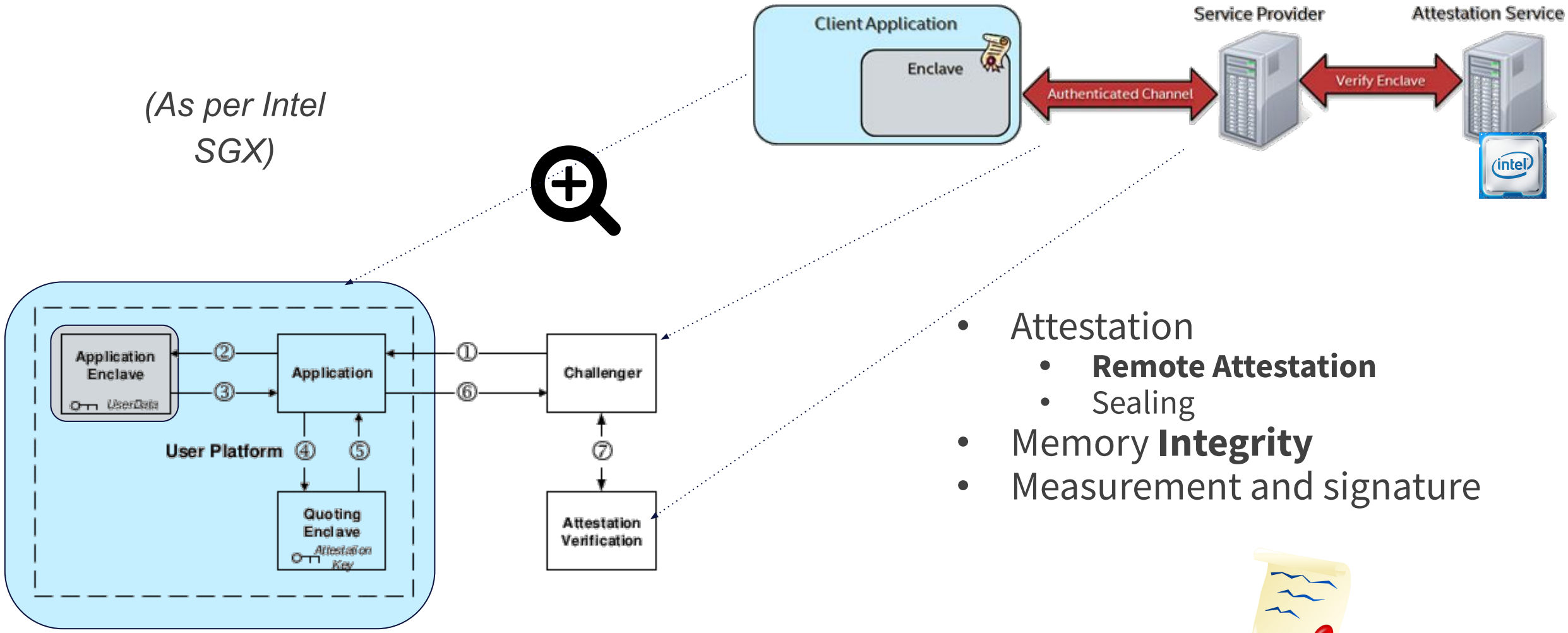


Confidential Computing BETA

Encrypt data in-use with Confidential VMs. Available in Beta for Google Compute Engine.

Under the Hood: Trusted Execution Environments

(As per Intel SGX)



- Attestation
  - **Remote Attestation**
  - Sealing
- Memory **Integrity**
- Measurement and signature

Secure Computation - Challenges

- Computational requirements
- Numeric stability and precision
- Trusted third parties / hardware vendors
- Adversaries and dishonest participants
- **Training is a challenge** on its own
- Leverage GPUs

Secure Inference - Performances

Library	Model X		Model Y		Model Z	
	Runtime (s)	Comm (MB)	Runtime (s)	Comm (MB)	Runtime (s)	Comm (MB)
PyTorch	0.0002	-	0.0003	-	0.0011	-
PySyft SNN	2.9	7.03	3.11	3.0	11.22	52.55
PySyft FSS	1.69	6.42	1.71	3.23	4.96	92.48
CrypTen	0.043	1.01	0.035	0.35	0.25	8.29
Graphene-SGX	0.009	-	0.04	-	0.053	-
Tensorflow	0.0003	-	0.0003	-	0.0008	-
TF-Trusted	0.14	-	0.12	-	0.14	-
TF-Encrypted	0.0112	-	0.012	-	0.132	-
HE-Transformer	15.4	1171.94	12.6	2780.66	548.94	56703.68

Table 1: Performance benchmarks during inference using a single data instance on MNIST [3] (Models X and Y) and the Malaria dataset [4] (Model Z).

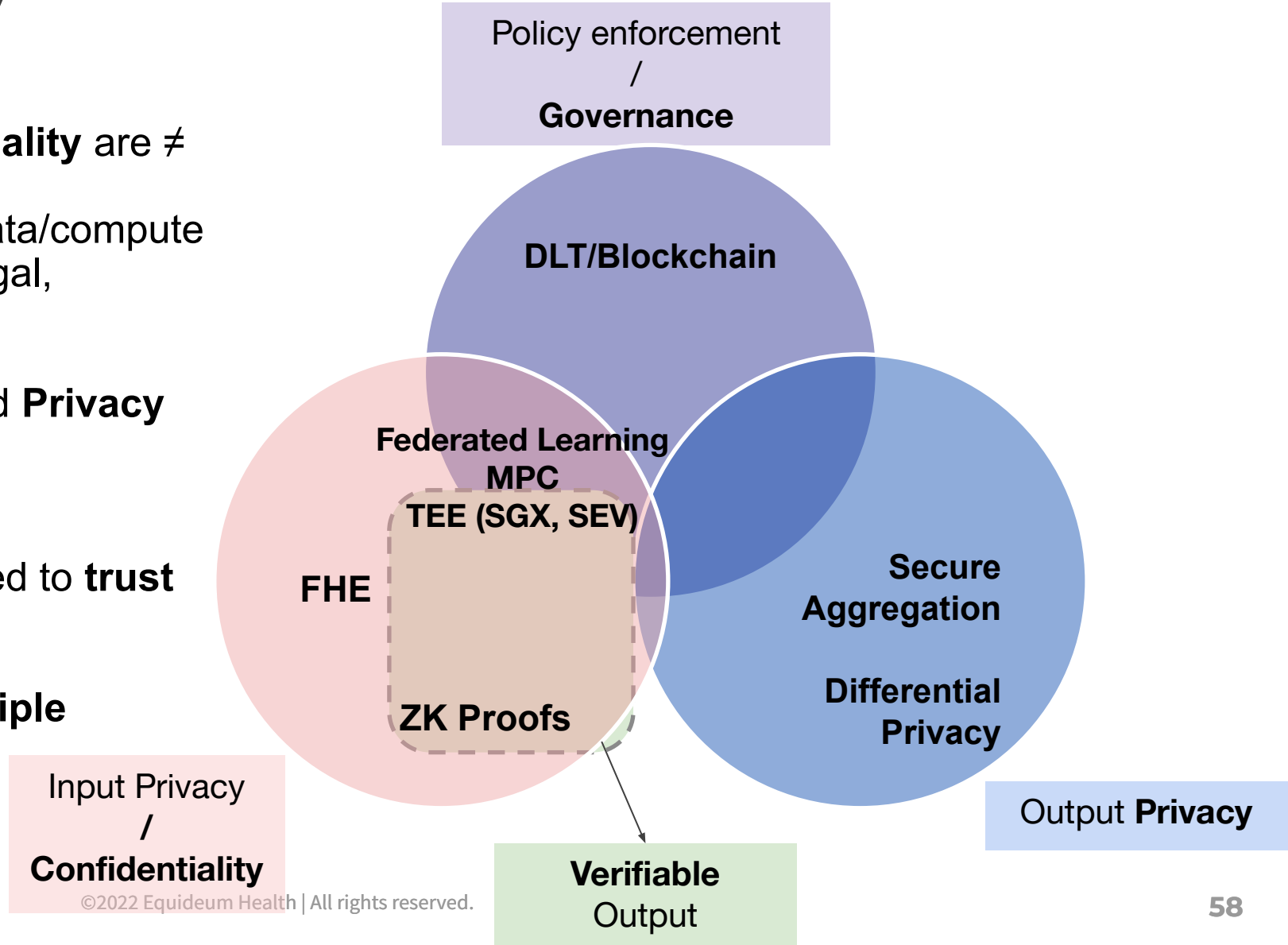
Haralampieva, V., *et al.* A Systematic Comparison of Encrypted Machine Learning Solutions for Image Classification. *PPMLP'20*
https://arxiv.org/abs/2011.05296

Secure Computation - Perspectives
The Compilers era

- Make low-level crypto primitives available to data-scientists
- **Automatic arithmetic circuit** generation
- Not only better UX, but also **parameters optimisation**
- Roster
  - Microsoft's EVA *(python-like)*
  - Zama's HNumpy
  - Google's C++ -> HE transpiler
  - Intel's HE Transformer *(tensorflow)*

Summary

- **Privacy, Verifiability, Confidentiality** are ≠
- Value fabricated by generating data/compute matches that were previously illegal, unethical and infeasible
- From an initial desire of increased **Privacy**
- Towards fostering **Collaboration**
- In a decentralised setting, we need to **trust results**
- PPML suggests **combining multiple PETs** to cover all aspects of Security and Privacy



References

- **Usynin, D., et al.** *Adversarial interference and its mitigations in privacy-preserving collaborative machine learning.* Nat Mach Intell 2021 https://doi.org/10.1038/s42256-021-00390-3
- **Usynin, D., et al.** *Zen and the art of model adaptation: Low-utility-cost attack mitigations in collaborative machine learning.* Proceedings on Privacy Enhancing Technologies, 2021 *(in press)*
- **Siomos et al.,** *Contribution Evaluation in Federated Learning: Examining Current Approaches*, New Frontiers in Federated Learning @NeurIPS, 2021
- **Kaissis et al,** *End-to-end privacy preserving deep learning on multi-institutional medical imaging*, 2021