A technical and responsible lens on

Privacy in machine learning

Afaf Taik

Content

1. What is privacy?
2. Differential privacy
3. Federated learning
4. Differentially private federated learning
5. Other techniques and open questions

What does privacy mean to you?

What does privacy mean to you?

Control what information about you is collected, used, or shared

What does privacy mean to you?

Control what information about you is collected, used, or shared

Protection of people's physical selves against invasive procedures

What does privacy mean to you?

Control what information about you is collected, used, or shared

Protection of people's physical selves against invasive procedures

Protection against unwarranted intrusion

What does privacy mean to you?

Control what information about you is collected, used, or shared

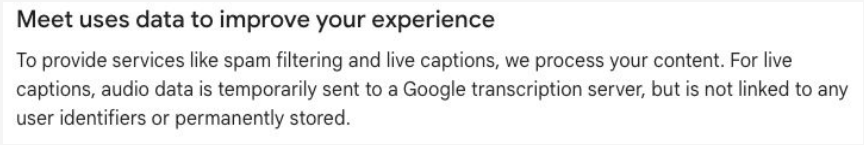
Protection of people's physical selves against invasive procedures

Protection against unwarranted intrusion

Protection of personal communication

Privacy in the world of Big Data

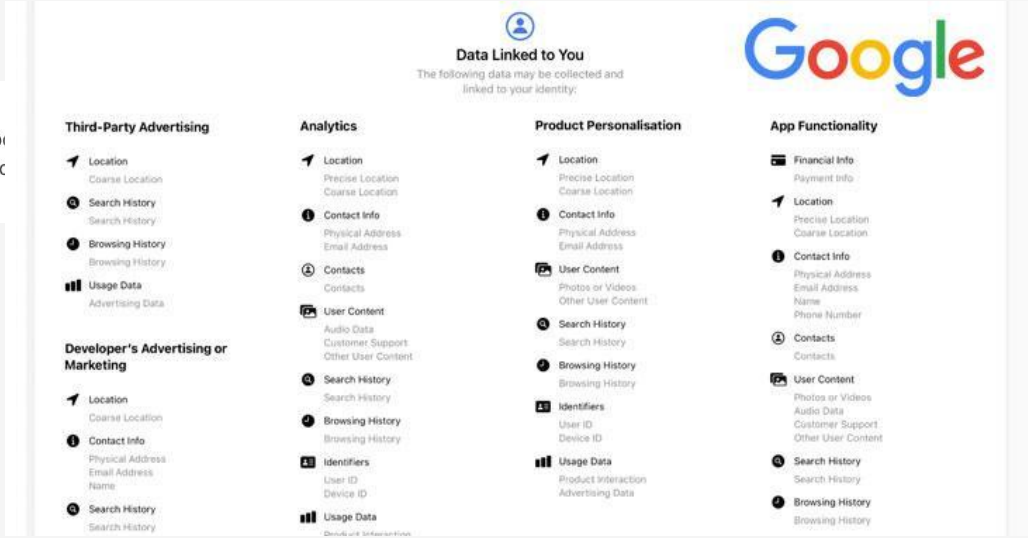
Privacy in the world of Big Data



Privacy in the world of Big Data

Meet uses data to improve your experience

To provide services like spam filtering and live captions, we pro
captions, audio data is temporarily sent to a Google transcriptic
user identifiers or permanently stored.



Privacy in the world of Big Data

Meet uses data

To provide services
captions, audio da
user identifiers or

Data Category	Description	Examples
Device connectivity and configuration data	This type of **Required diagnostic data** includes details about the device, its configuration, and connectivity capabilities.	• Device properties such as the OEM manufacturer, processor type, and memory attributes. • Device settings and configurations, such as networking and peripherals data.
Product and service performance data	This type of **Required diagnostic data** includes details about device or service health and performance.	• Basic error reporting, such as whether updates were successfully installed. • Reliability data about the health of the operating system or services.
Software setup and inventory data	This type of **Required diagnostic data** includes software installation and update information on the device.	• Operating system version, configuration details and updates installed. • Apps and drivers installed on the device.

Privacy in the world of Big Data

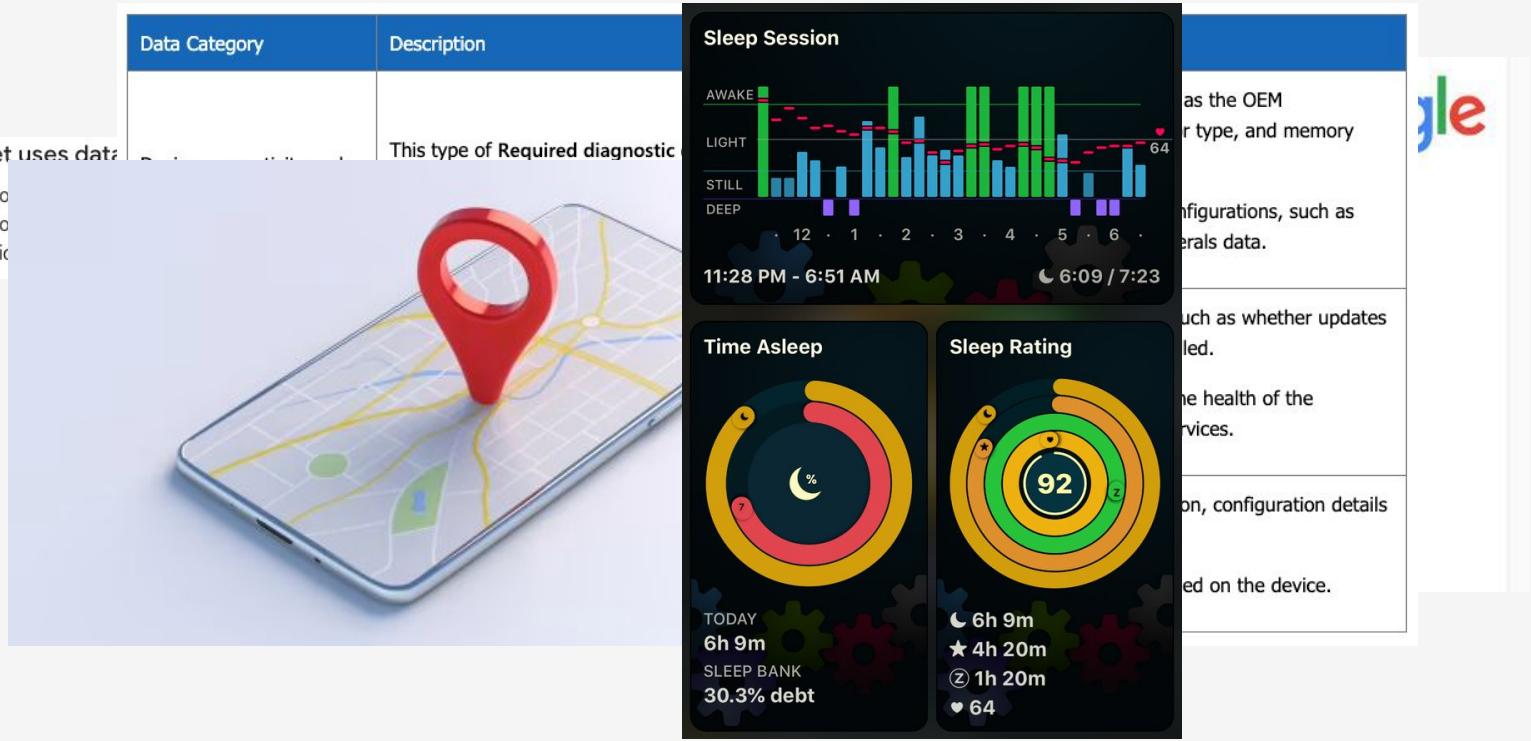
Meet uses data

To pro
captio
user id

Data Category	Description	Examples
	This type of **Required diagnostic data** includes details	• Device properties such as the OEM manufacturer, processor type, and memory attributes. Device settings and configurations, such as networking and peripherals data.
		Basic error reporting, such as whether updates were successfully installed. Reliability data about the health of the operating system or services.
		Operating system version, configuration details and updates installed. Apps and drivers installed on the device.

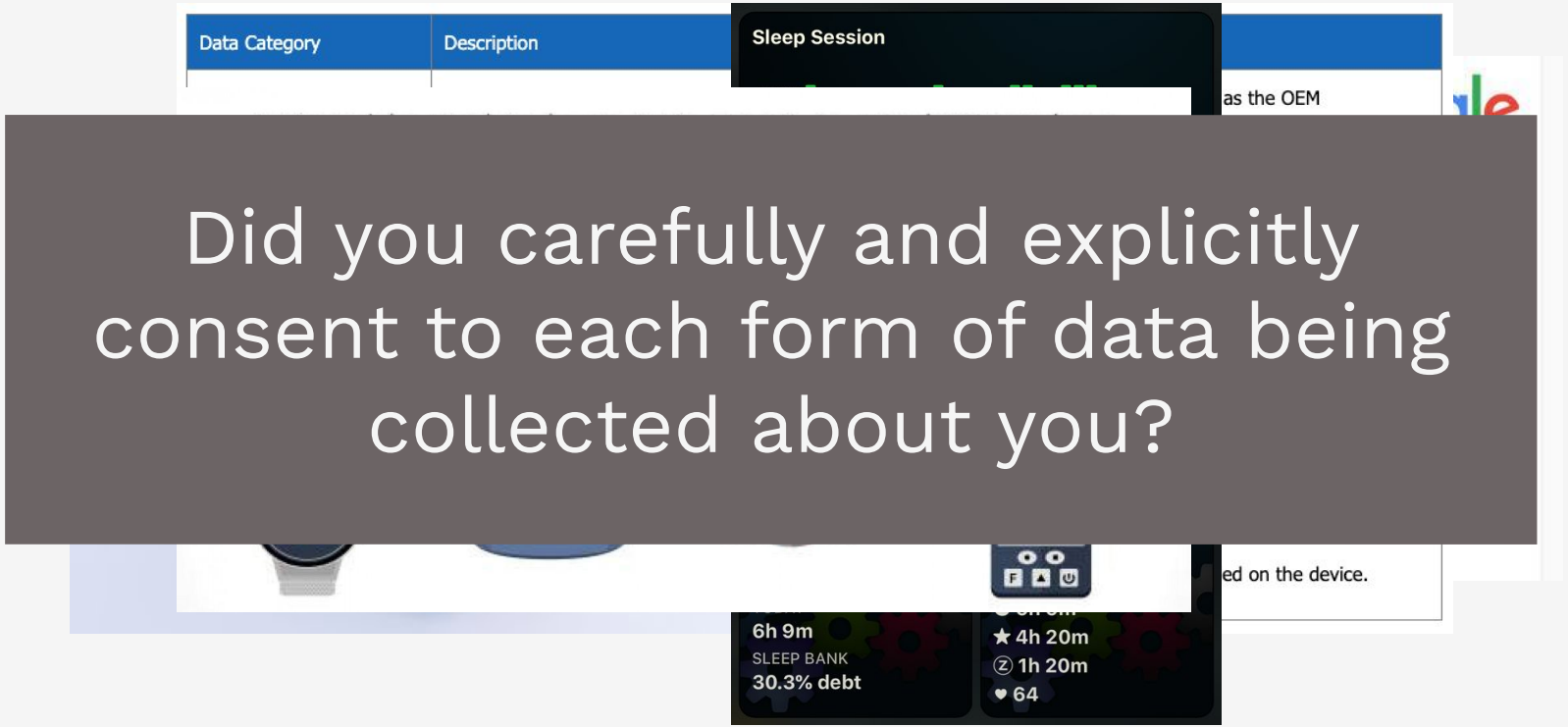
gle

Privacy in the world of Big Data



Privacy in the world of Big Data

Privacy in the world of Big Data



Did you carefully and explicitly consent to each form of data being collected about you?

Privacy in the world of Big Data

Data Privacy: “*relationship between the collection and dissemination of data, technology, the public expectation of privacy, contextual information norms, and the legal and political issues surrounding them*”

Privacy in the world of Big Data

Data Privacy: "*relationship between the collection and dissemination of data, technology, the public expectation of privacy, contextual information norms, and the legal and political issues surrounding them*"

But why care?

Privacy in the world of Big Data

Cambridge Analytica and Facebook: The Scandal and the Fallout So Far

Revelations that digital consultants to the Trump campaign misused the data of millions of Facebook users set off a furor on both sides of the Atlantic. This is how The Times covered it.

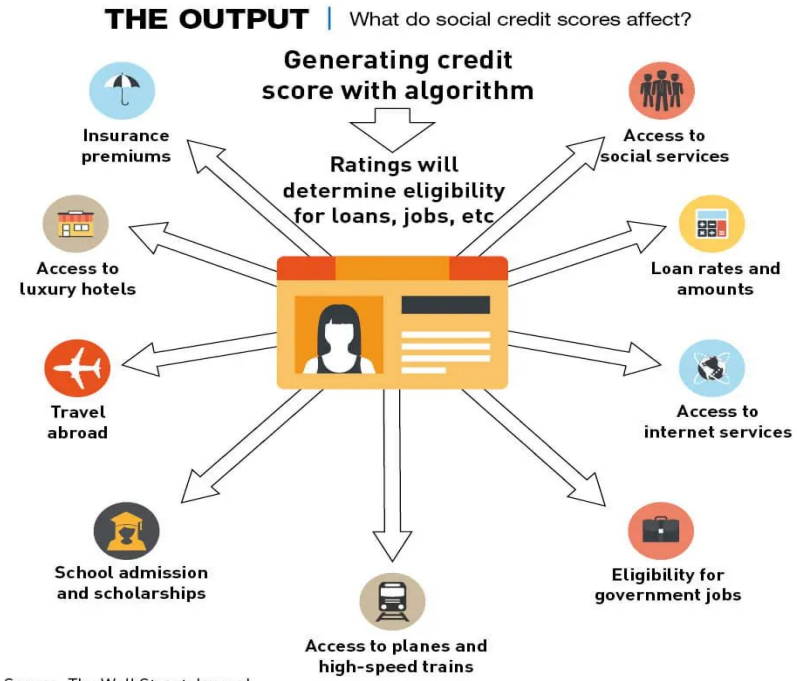
Privacy in the world of Big Data

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

- Target discovered that
  - Pregnant women purchased large quantities of unscented lotion at the beginning of the 2nd trimester.
  - They purchased supplements of calcium, magnesium, and zinc for the first 20 weeks.
  - They purchased scent-free soap and extra-large bags of cotton balls.

  - In total, Target identified 25 items that could act as pregnancy predictors.

Privacy in the world of Big Data



Privacy in the world of Big Data

When you don’t have control over how your data is being used, you don’t have control over how it can affect you!

Data Protection Regulations

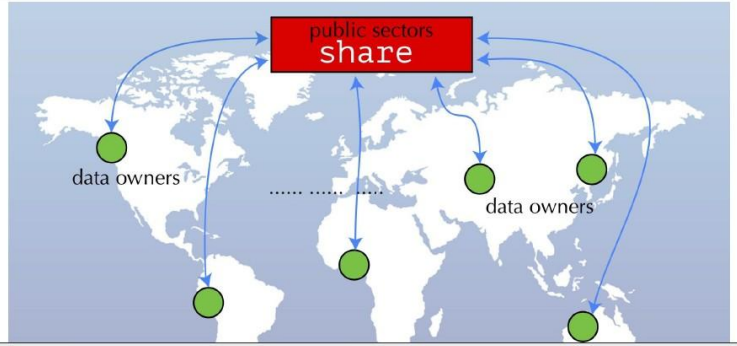
Data Protection Regulations

2020	2018	2022	2023
			
Brazil LGPD	EU GDPR	Dubai PDPA	Colorado CPA
2021	2023	2020	2022
			
Virginia VDPA	Connecticut DPA	South African POPIA	Thailand PDPA


Punishable by a fine
#StarTrekLawyerPosting
Means legal for a price

Data philanthropy

[United Nations Global Pulse]



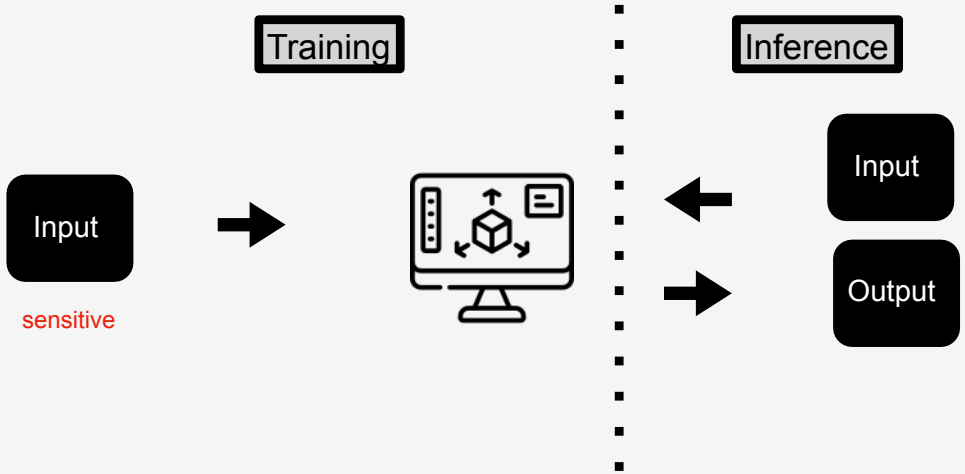
"--- think of big **data** as a new kind of natural resource – **infinitely renewable, increasingly ubiquitous** – but … has fallen into the hands of… industry …**Data has a social opportunity** – and we have **a social responsibility** – …***data reaches the people who need it most***."

[Director of UN Global Pulse]

Your data is important for public good

So how can we achieve privacy?

Privacy Setting in ML



- … learning nothing about an individual while learning useful information about a population…

Privacy by information control

- Conventional measures for privacy:
  - Control access to information
  - Control the flow of information
  - Control the purposes information is used

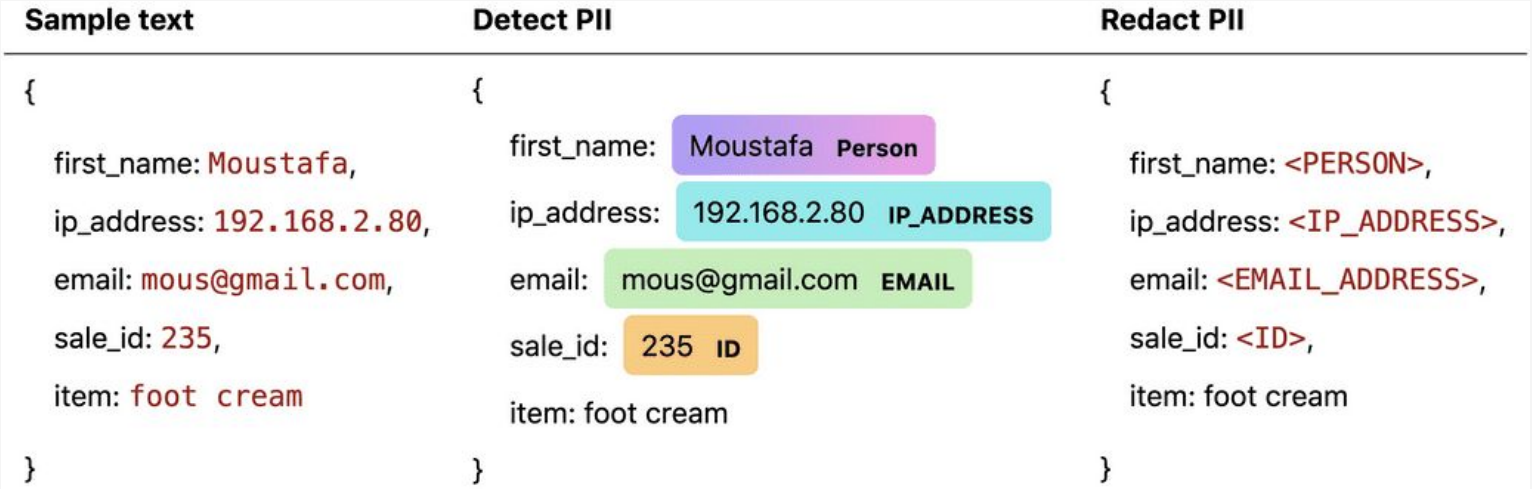
Privacy by information control

- Conventional measures for privacy:
  - Control access to information
  - Control the flow of information
  - Control the purposes information is used

They basically do not guarantee privacy!

Privacy via anonymization

the process of protecting private or sensitive information by erasing or encrypting identifiers that connect an individual to stored data.



New York taxi details can be extracted from anonymised data, researchers say

FoI request reveals data on 173m individual trips in US city - but could yield more details, such as drivers' addresses and income

Privacy via anonymization

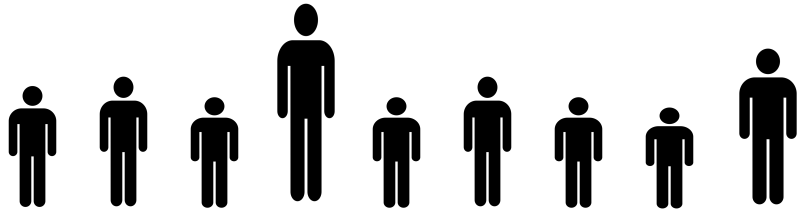
WIRED SECURITY POLITICS GEAR THE BIG STORY BUSINESS SCIENCE CULTURE IDEAS MERCH SIGN IN SUBSCRIBE

RYAN SINGEL SECURITY MAR 12, 2010 2:48 PM

NetFlix Cancels Recommendation Contest After Privacy Lawsuit

Netflix is canceling its second $1 million Netflix Prize to settle a legal challenge that it breached customer privacy as part of the first contest's race for a better movie-recommendation engine. Friday's announcement came five months after Netflix had announced a successor to its algorithm-improvement contest. The company at the time said it intended to [...]

What about aggregate data?



Bob

Group A

Average Height = 5’10

Group B

Average Height = 5’6

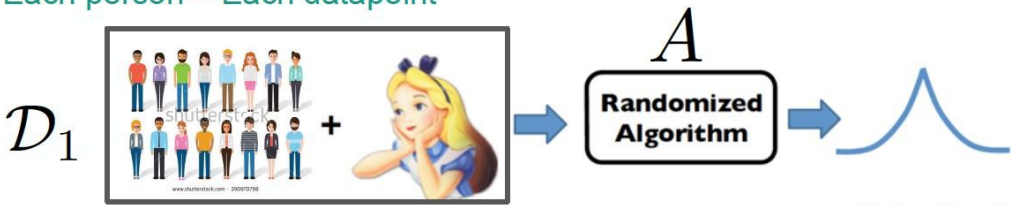
Can you guess which group Bob belongs to?

Differential privacy

Differential Privacy

[Dwork 06] -

Each person = Each datapoint

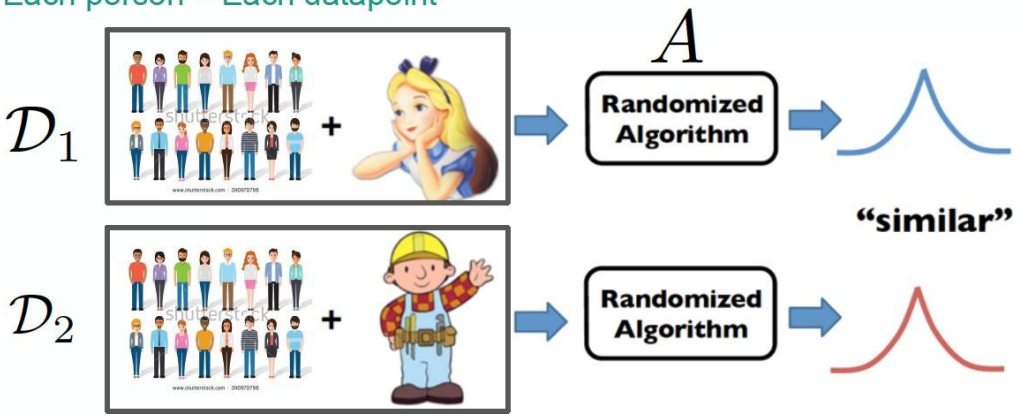


Differential privacy is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any individual in it. (Greenberg, 2016)

Differential Privacy

[Dwork 06] -

Each person = Each datapoint



Differential privacy is the statistical science of trying to learn as much as possible about a group while learning as little as possible about any individual in it. (Greenberg, 2016)

Formal Definition of ε-DP

A randomized algorithm **A** satisfies **ϵ-differential privacy** if, for any two datasets **D** and **D′** that differ by only one individual's data and for any possible output **O** of the algorithm:

$$\Pr[\mathcal{A}(D) = O] \leq e^{\epsilon} \cdot \Pr[\mathcal{A}(D') = O]$$

Here, ϵ controls how much the inclusion of a single data point affects the model’s output.

What is ε?

The privacy parameter ϵ quantifies the privacy loss: **smaller ϵ** values provide **stronger privacy.**

A typical ϵ value might range from 0.01 to 10 depending on the use case.

In practice, ϵ is chosen based on the trade-off between privacy and accuracy—lower ϵ gives better privacy but can reduce utility.

Differential Privacy

Typically add randomness (noise) for DP

How much noise to add?

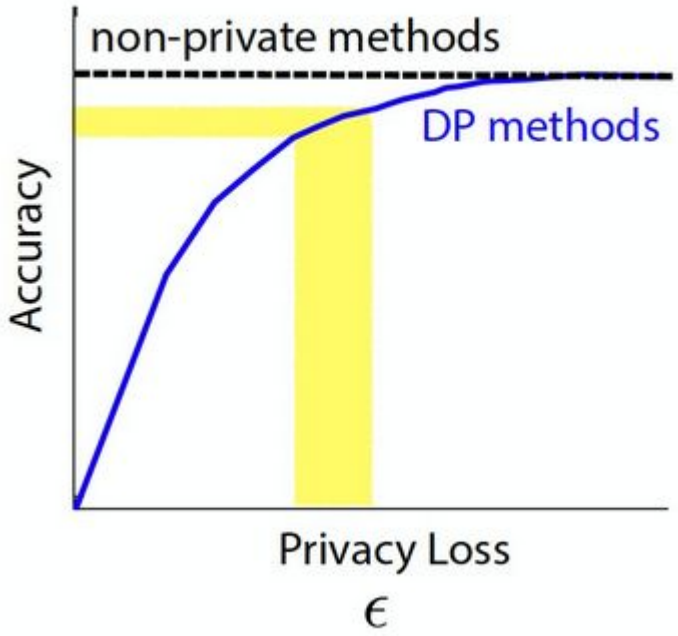
Sensitivity (Δf)

Sensitivity Δf is the maximum amount a function f can change when one data point in the dataset changes.

$$\Delta f = \max_{D,D'} |f(D) - f(D')|$$

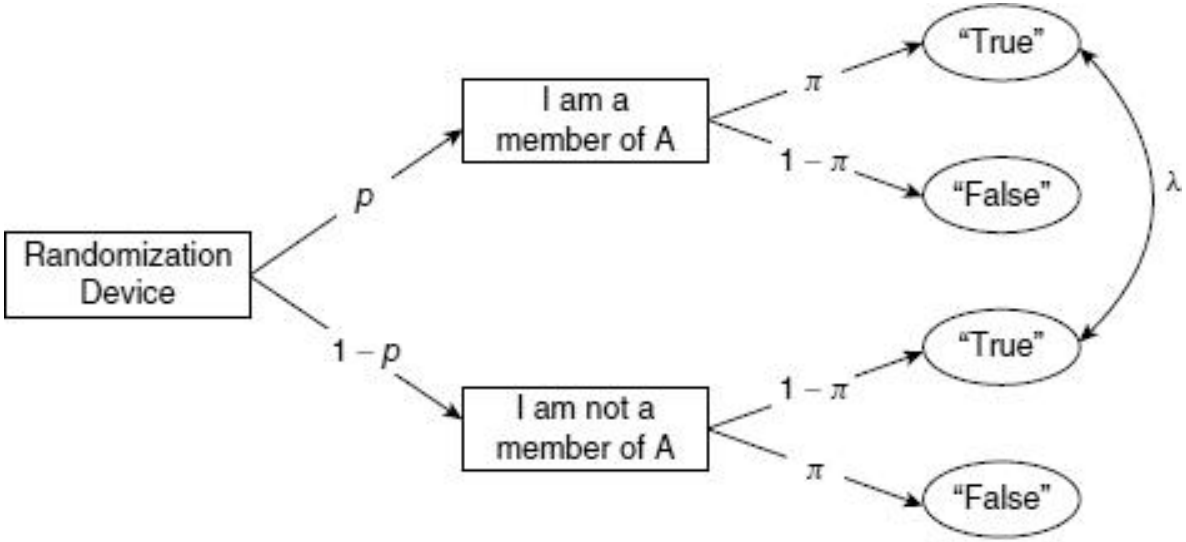
$$\text{noise} \propto \frac{\text{sensitivity}}{\text{privacy loss}}$$

Privacy & Accuracy tradeoff



Privacy mechanisms

Randomized response



$$\lambda = p\pi + (1 - p)(1 - \pi)$$

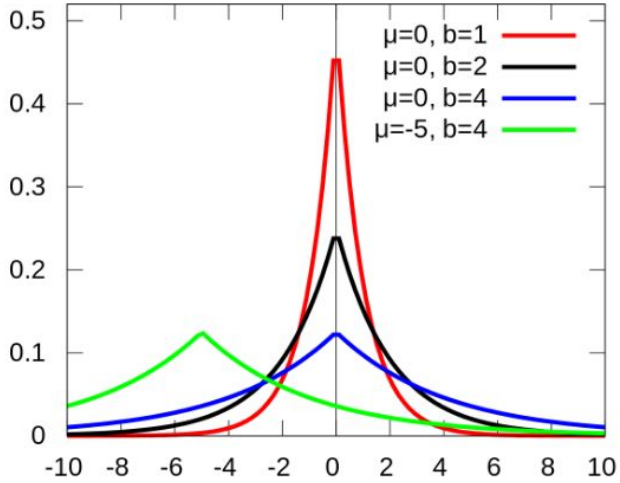
Laplace mechanism

Laplace density function:

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

$$\mathcal{M}_{\text{Lap}}(x, f, \epsilon) = f(x) + \text{Lap}\left(\mu = 0, b = \frac{\Delta f}{\epsilon}\right)$$

Source: Wikipedia



Laplace distributions with various parameters.

Laplace mechanism is ε-DP

$$
\begin{aligned}
\frac{\Pr(\mathcal{M}_{\text{Lap}}(x, f, \epsilon) = z)}{\Pr(\mathcal{M}_{\text{Lap}}(y, f, \epsilon) = z)} &= \frac{\Pr(f(x) + \text{Lap}(0, \frac{\Delta f}{\epsilon}) = z)}{\Pr(f(y) + \text{Lap}(0, \frac{\Delta f}{\epsilon}) = z)} \\
&= \frac{\Pr(\text{Lap}(0, \frac{\Delta f}{\epsilon}) = z - f(x))}{\Pr(\text{Lap}(0, \frac{\Delta f}{\epsilon}) = z - f(y))} \\
&= \frac{\frac{1}{2b}\exp\left(-\frac{|z-f(x)|}{b}\right)}{\frac{1}{2b}\exp\left(-\frac{|z-f(y)|}{b}\right)} \\
&= \exp\left(\frac{|z - f(y)| - |z - f(x)|}{b}\right) \\
&\leq \exp\left(\frac{|f(y) - f(x)|}{b}\right) \\
&\leq \exp\left(\frac{\Delta f}{b}\right) = \exp(\epsilon).
\end{aligned}
$$

$$
\mathcal{M}_{\text{Lap}}(x, f, \epsilon) = f(x) + \text{Lap}\left(\mu = 0, b = \frac{\Delta f}{\epsilon}\right)
$$

Gaussian mechanism (ε, δ)-DP

Given a function f with L2 sensitivity Δf, the Gaussian Mechanism perturbs f(D) by adding noise drawn from a Gaussian distribution with mean 0 and standard deviation σ proportional to Δf.

$$\mathcal{M}(D) = f(D) + \mathcal{N}(0, \sigma^2)$$

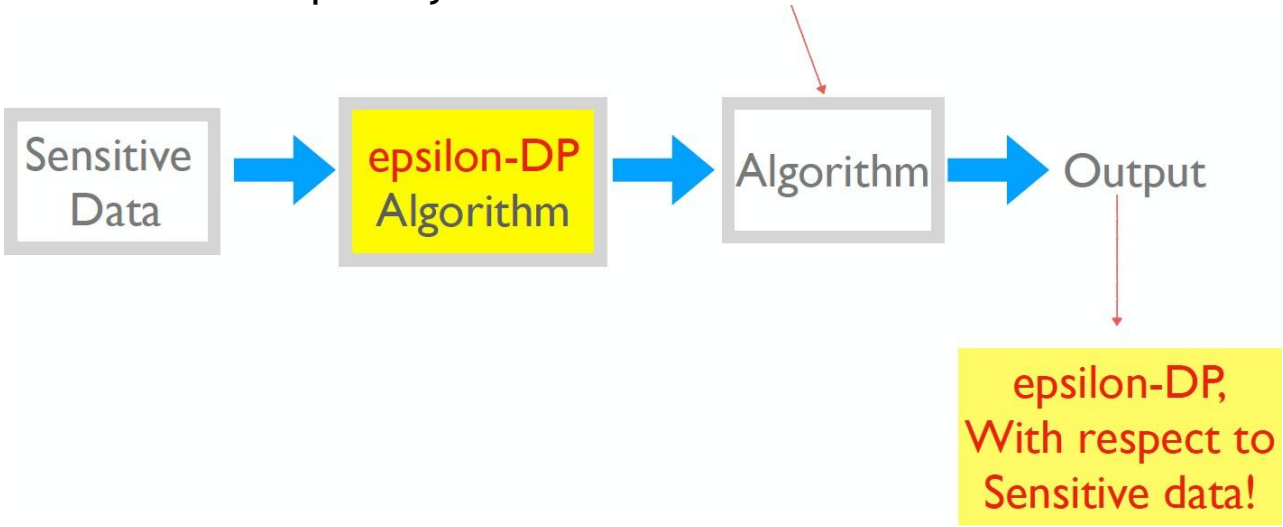
Laplace mechanism gives exact guarantees while gaussian gives approximate privacy guarantees.

$$\Pr[\mathcal{M}(D) = O] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(D') = O] + \delta$$

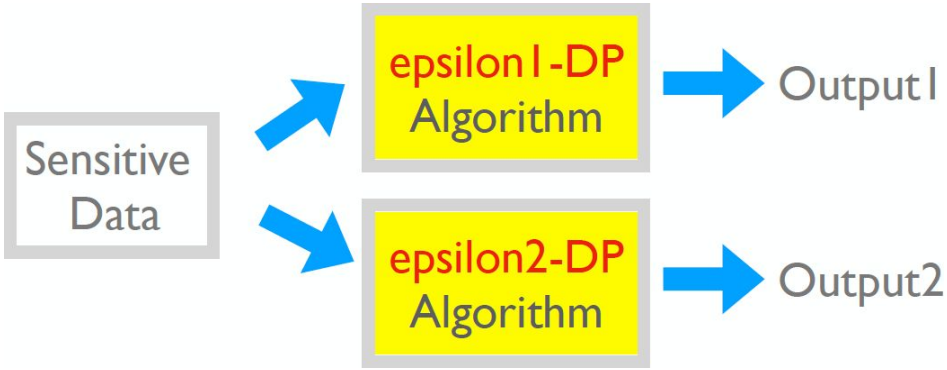
Properties of DP

Post-processing invariance

- Differential privacy is immune to



Composability



Union of output 1 & output 2 is (epsilon 1+epsilon 2)-DP!

Privacy News

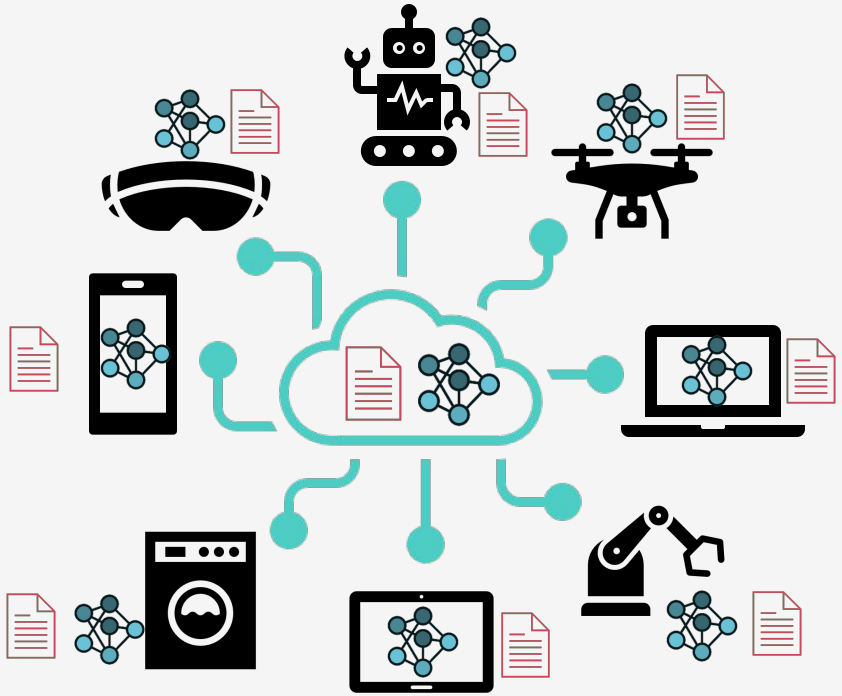
Big tech doesn't care about your data: It's just "privacy washing"!

As people around the world become more aware of the importance of protecting their data & privacy, big tech corporations like Google and Apple adapt to appear privacy focused; welcome to "privacy washing"!

epsilon—smaller epsilon means more private. Theoreticians generally agree that an epsilon less than 0.1 is very safe, and an epsilon less than 1.0 is probably ok. The USC, Indiana, and Tsinghua researchers reveal that Apple's MacOS implementation uses an epsilon of 6, while iOS 10 uses an epsilon of 14. So what does this tell us about how private Apple's data collection is? Or to take a concrete example, if the government were to subpoena

Federated Learning

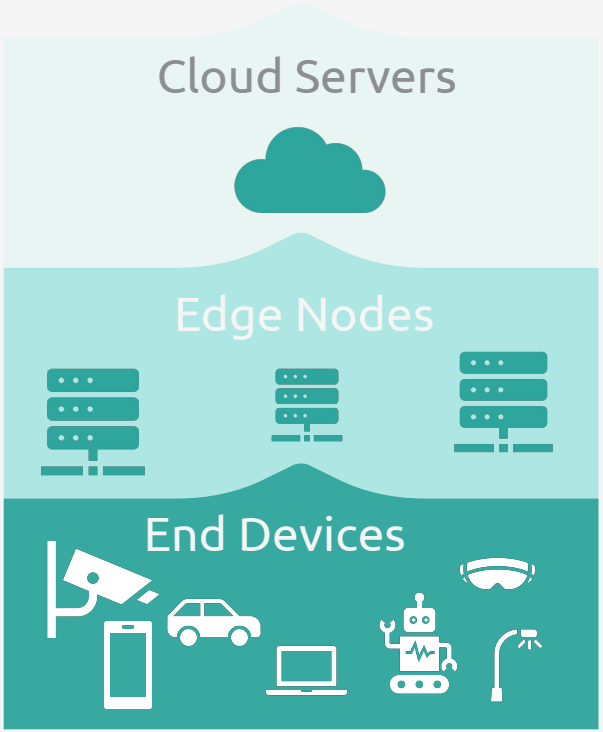
And edge computing I guess!

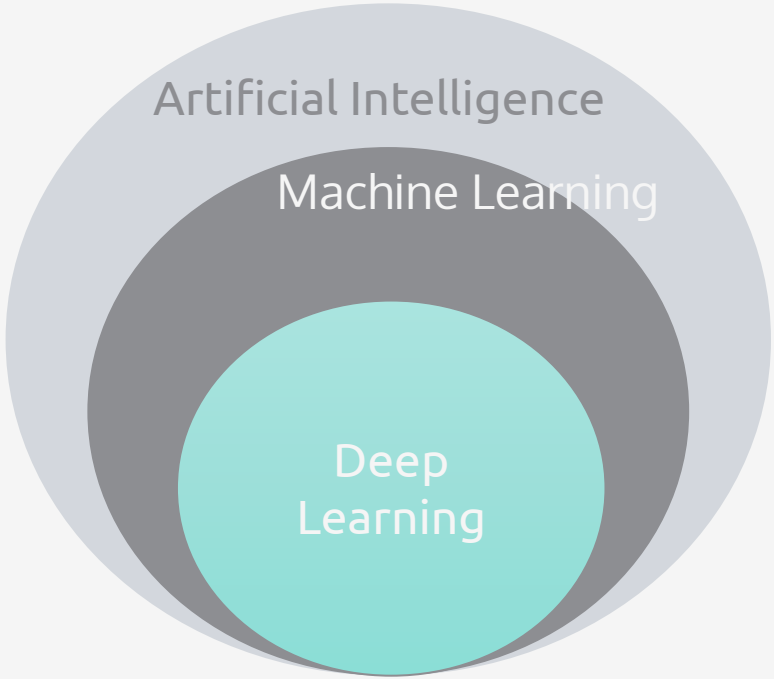


- Communication overhead
- High energy consumption
- Privacy concerns

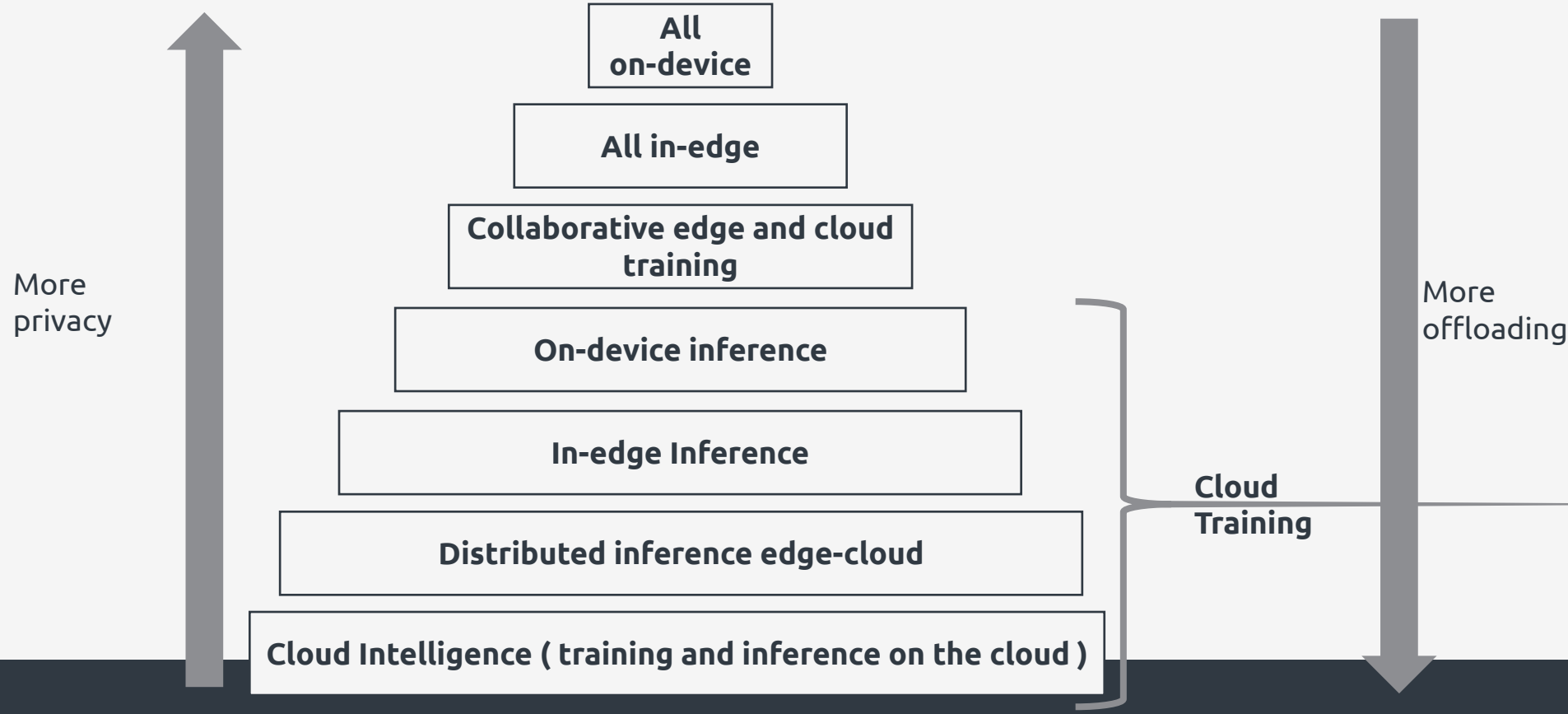
How can we address these challenges?

Edge Intelligence





Edge Intelligence Deployment

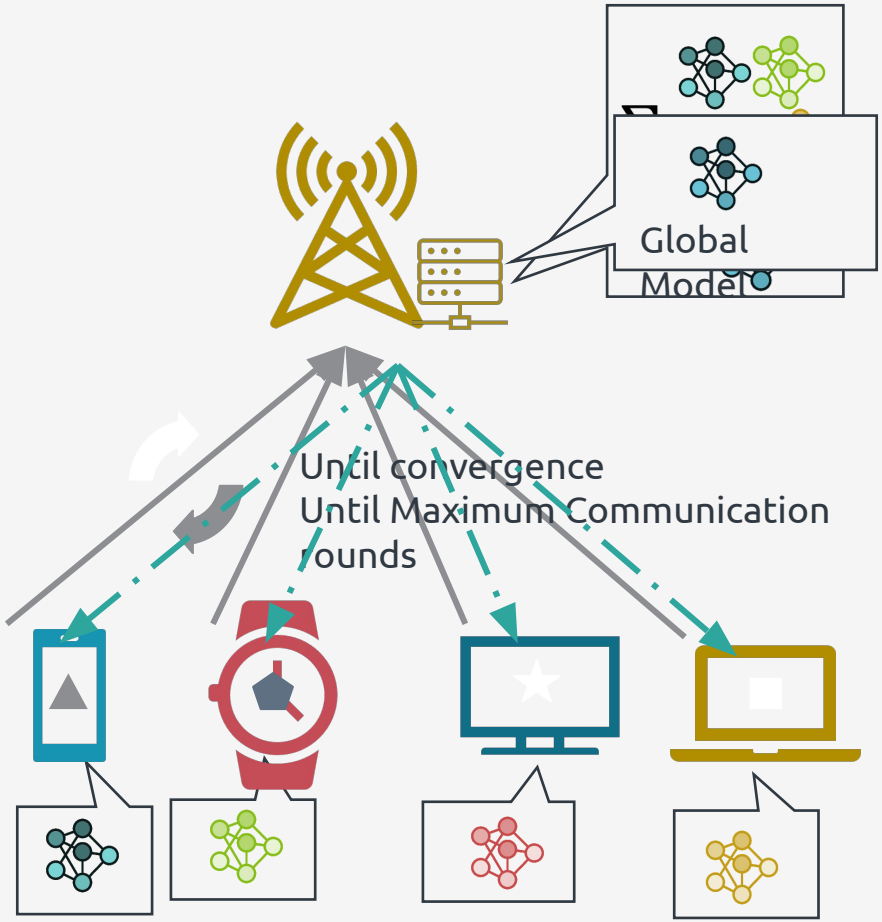


FL Terminology

Clients :individual devices or entities that hold private, locally-stored data. Examples of clients include smartphones, IoT devices, hospitals, or other organizations.

Server: a central entity responsible for coordinating the learning process across multiple clients.

Federated Learning



Applications of Federated

What makes a good application?

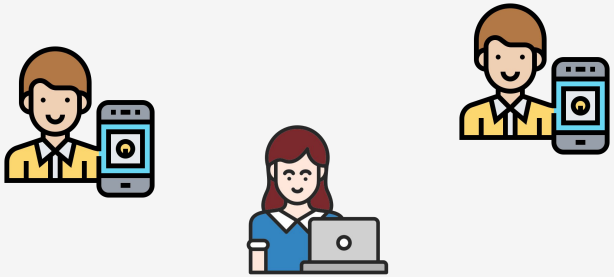
- On-device data is more relevant than server-side proxy data
- On-device data is privacy sensitive or large
- Labels can be inferred naturally from user interaction

Example applications

- Language modeling for mobile keyboards and voice recognition
- Image classification for predicting which photos people will share
- ...

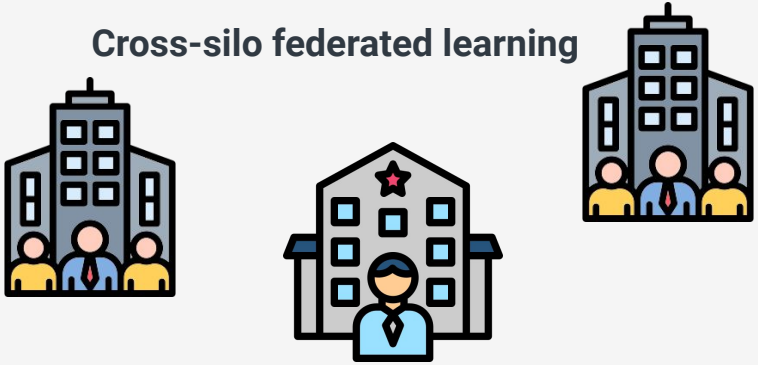
Variations of Federated Learning

Cross-device federated learning



Mobile phones, IoT devices, voice assistants, activity trackers, smart appliances…

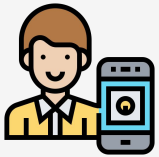
Cross-silo federated learning



Banks, hospitals, schools,...

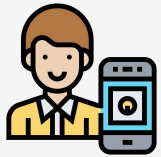
Variations of Federated Learning

Horizontal federated learning

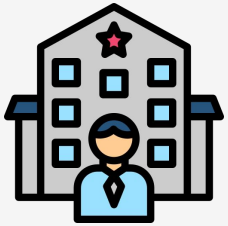


X1, X2, Y

X1, X2, Y



X1, X2, Y



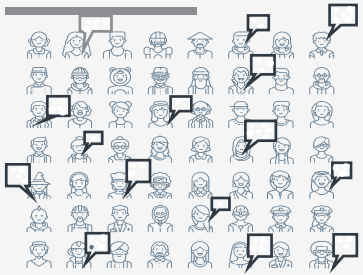
X3, Y

Vertical federated learning

X1, X2

Data-related challenges

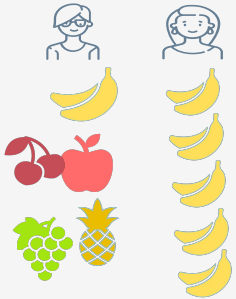
Small Datasets



Non- i.i.d

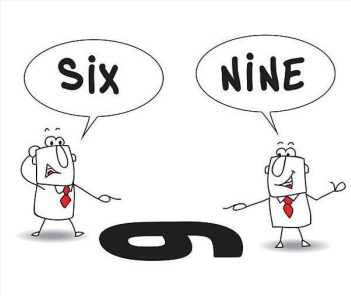
Imbalance

Redundancy



Reliability

Concept-shift



Edge environment challenges

Heterogeneity

Hardware

Memory

Connectivity

Battery

Scarcity

Bandwidth

Time

Available devices

Mobility

Limited time

Uncertainty

Solutions and optimization axis

- Client selection and resource allocation: Optimizing who participates and allocating resources to achieve a certain participation goal
- Local training: structured updates
- Updates upload: sketched updates, quantization, compression
- Aggregation: synchronous vs asynchronous, different weighting schemes
- Personalization : local retraining, clustered FL, multi-task FL

Is this enough?

Information Leakage through ML Models

Generalization and Memorization

Generalization and Memorization

barn swallow

tree swallow

Generalization and Memorization

Can you tell me if this is a barn swallow or a tree swallow?

Generalization and Memorization

barn swallow

tree swallow

Can you tell me if this is a barn swallow or a tree swallow?

Generalization and Memorization

barn swallow

tree swallow

Generalization and Memorization

Can you tell me if this is a barn swallow or a tree swallow?

Generalization and Memorization

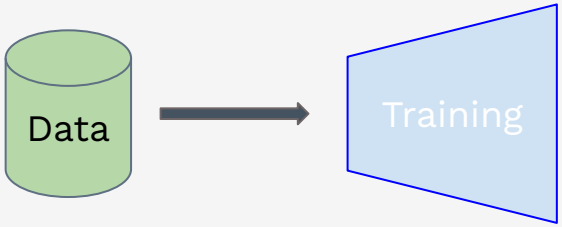
Can you tell me if this is a barn swallow or a tree swallow?

Can you tell me if this is a barn swallow or a tree swallow?

Generalization and Memorization

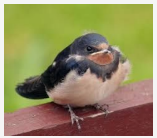
Generalization: *the ability to perform well on unseen data.*

Information Leakage through ML Models

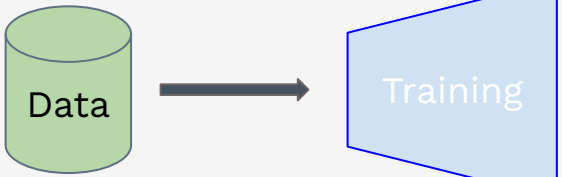


Barn swallows have cinnamon colored underparts while tree swallows have white underparts.

Was this image part of your data?



Can’t tell



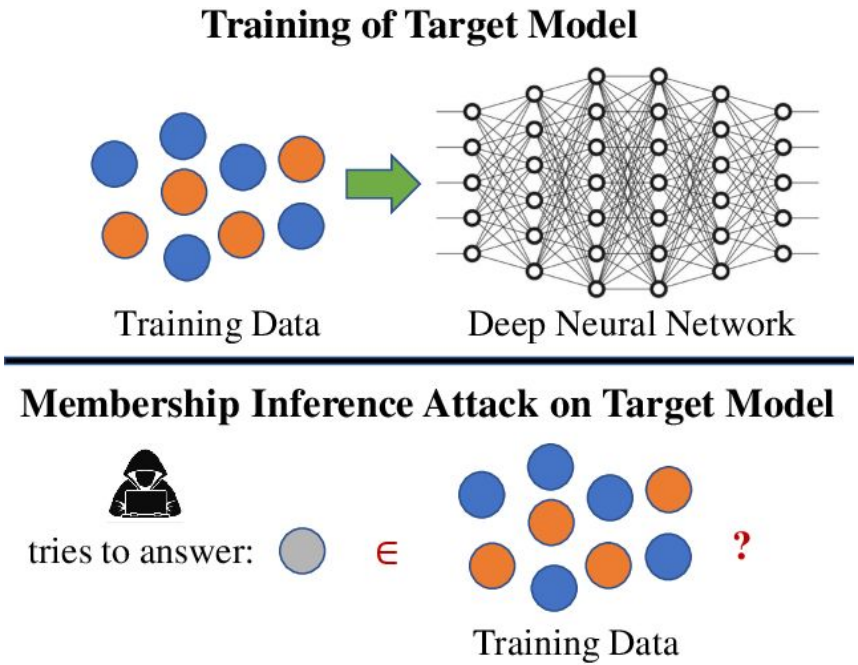
Memorize the photos!

Was this image part of your data?

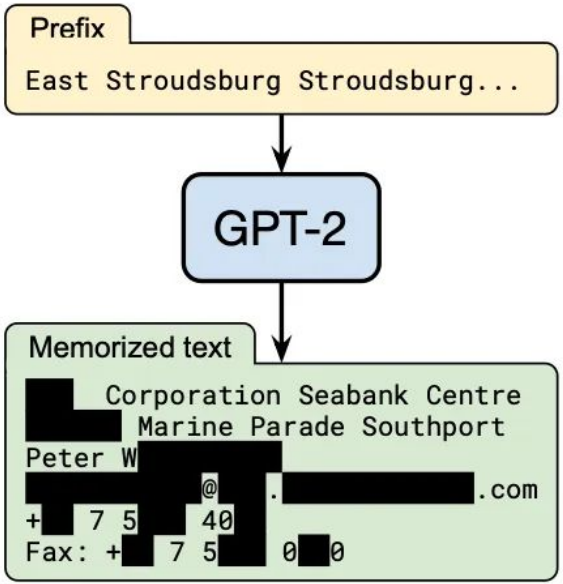
Yes it was

Information Leakage through ML Models

Membership Inference Attacks



Data Extraction Attacks



Model Extraction Attacks

De-Anonymization Attacks

Property Inference Attacks

Side-Channel Attacks

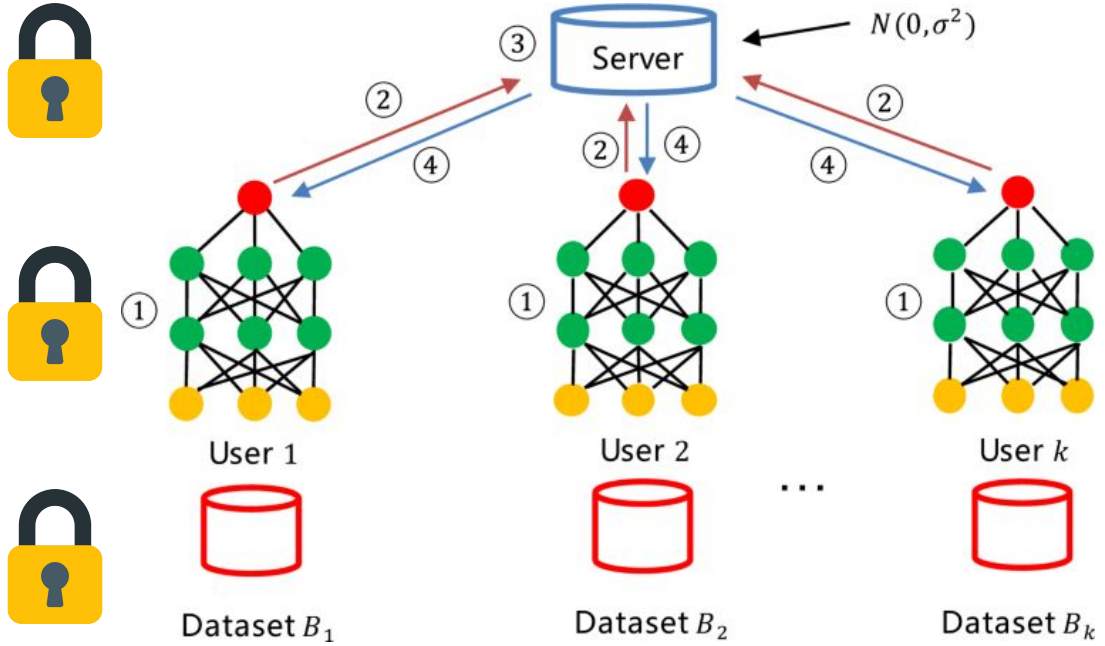
.....

Differentially private federated learning

How much trust is there?

Trust the server
Trust the communication channel

Don’t trust the server and/or
Communication channel



DP-SGD

Key Steps of DP-SGD:

1. **Gradient Computation**: Just like in regular SGD, the gradients of the loss function are computed for a batch of data.
2. **Gradient Clipping**: This is done to limit the influence of any single data point on the model's updates, which is essential for privacy.
3. **Adding Noise**: After clipping the gradients, noise is added to the sum of the gradients (typically sampled from a Gaussian distribution) with the scale of the noise determined by a parameter σ (the noise multiplier).
4. **Updating the Model**: The model is updated using the noisy, clipped gradients. This ensures that even if someone inspects the gradient updates, they cannot learn much about any individual data point.

Abadi et al. (2016)

Importance of Clipping

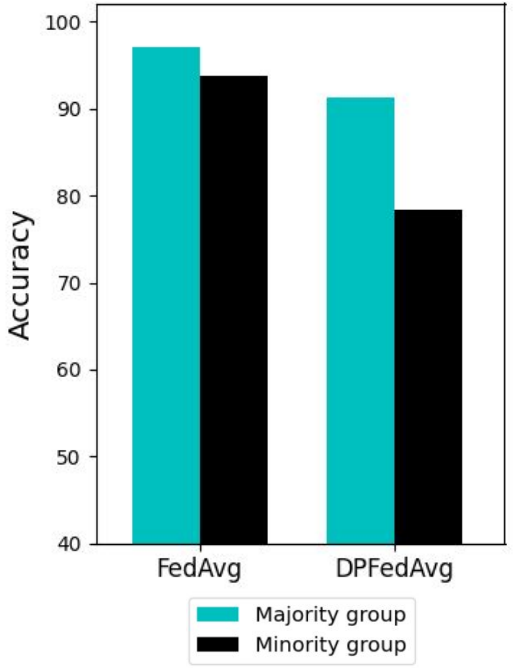
Privacy: Lowering the clipping threshold C reduces the sensitivity of the gradients, allowing less noise to be added for a given privacy level ϵ. However, if C is too low, it can lead to excessive gradient clipping, degrading model accuracy.

Utility: Setting C higher preserves more of the natural gradient information, which helps in learning and improves model accuracy. But this also increases sensitivity, requiring more noise to achieve the same level of privacy, potentially impacting performance.

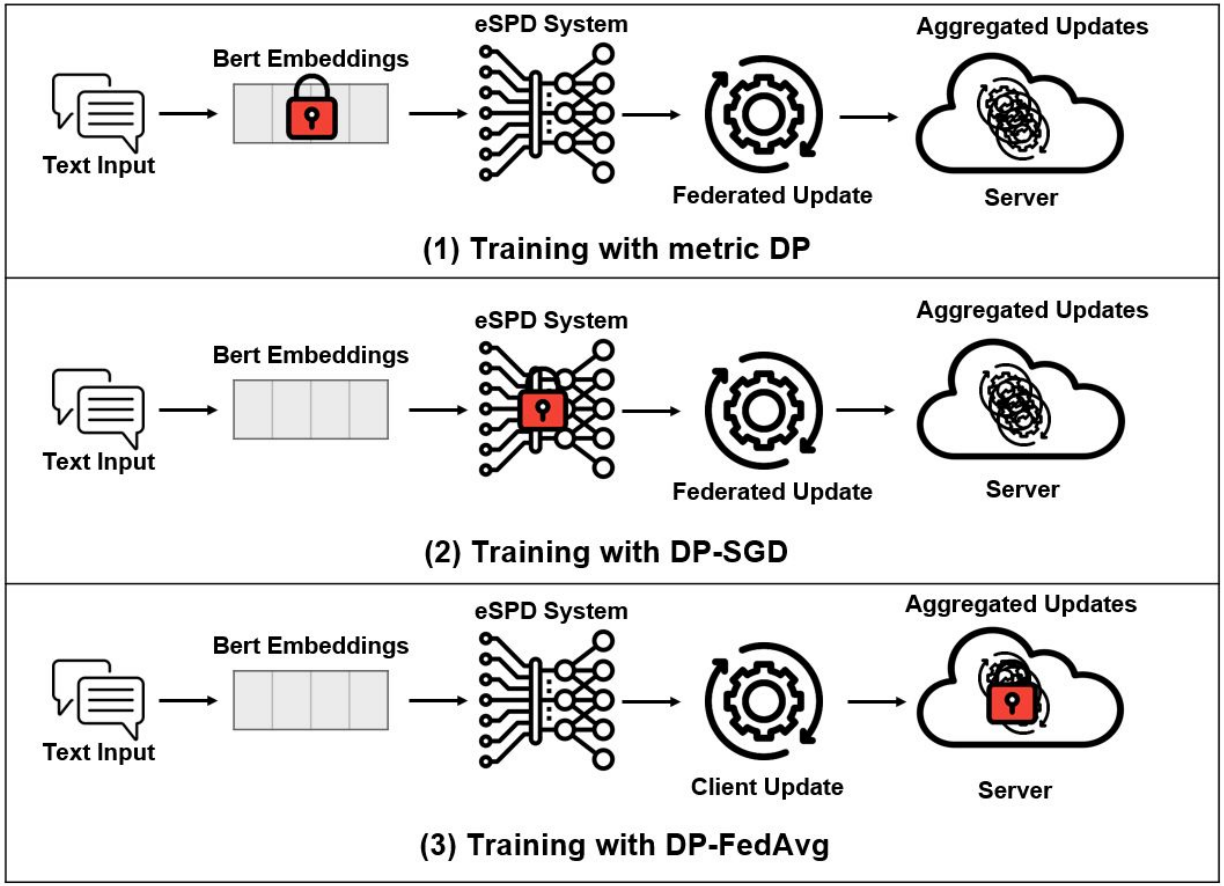
But is DPFL a perfect solution?

DPFL exacerbates data heterogeneity issues

For some clients, privacy comes at a higher cost compared to others!



How much trust is there?



Use case and Dilemma

Privacy for whom?

If E2EE is rolled out widely without necessary child safety measures, social media companies will no longer be able to find and report child sexual abuse material in the same way.

Google refuses to reinstate man's account after he took medical images of son's groin

Experts say case highlights dangers of automated detection of child sexual abuse images

Other privacy related topics

Data Minimization

1. Personal data shall be:

   (a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

   (b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');

   (c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

Right to be Forgotten

Art. 17 GDPR

Right to erasure ('right to be forgotten')

1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:

   a. the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;

   b. the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;

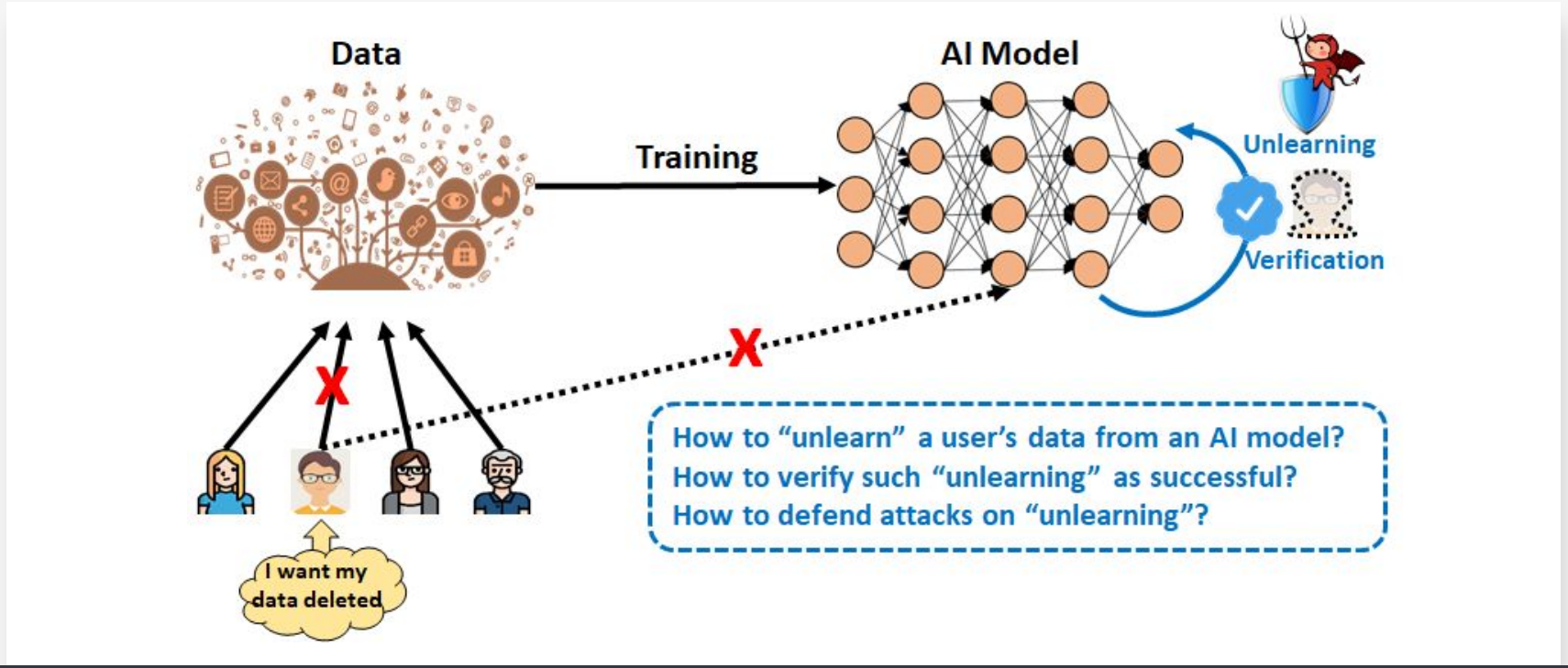
   c. the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);

   d. the personal data have been unlawfully processed;

   e. the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;

   f. the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).

Right to be Forgotten



Libraries and frameworks

To conclude

Privacy huh!

Such broad topic…

As a practitioner:
- Remember to think about how you are collecting, processing, storing and sharing data and models.

As a user:
Be careful what you agree to and opt out of unnecessary data collection!

Acknowledgement

Part of the credit for these slides and figures and work goes to:

Prakhar Ganesh

Golnoosh Farnadi

Data Minimization

Meta Fined $1.3 Billion for Violating E.U. Data Privacy Rules

South Korea's PIPC fines Meta for exceeding data minimization standards