Mila

23-07-2024

Data Privacy in Machine Learning

Prakhar Ganesh

Before we start …

How’s everyone doing?

Final lecture! Excited?!

Goals today...

- Introduce Privacy in Machine Learning
- Understand the relationship between Memorization and Generalization
- Introduce Anonymization, Differential Privacy, and other privacy frameworks
- Get familiar with the legal frameworks of Privacy

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

What does privacy mean to you?

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

P

a

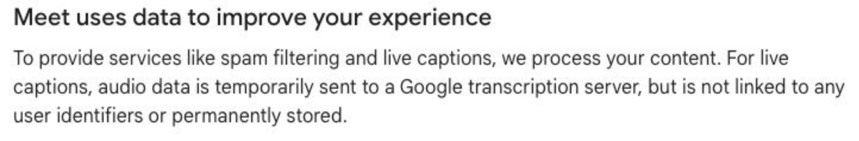
The Right to be Left Alone

Protection against unwarranted intrusion

Protection of personal communication

Privacy in the world of Big Data

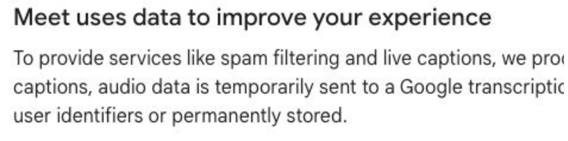
Privacy in the world of Big Data



Meet uses data to improve your experience

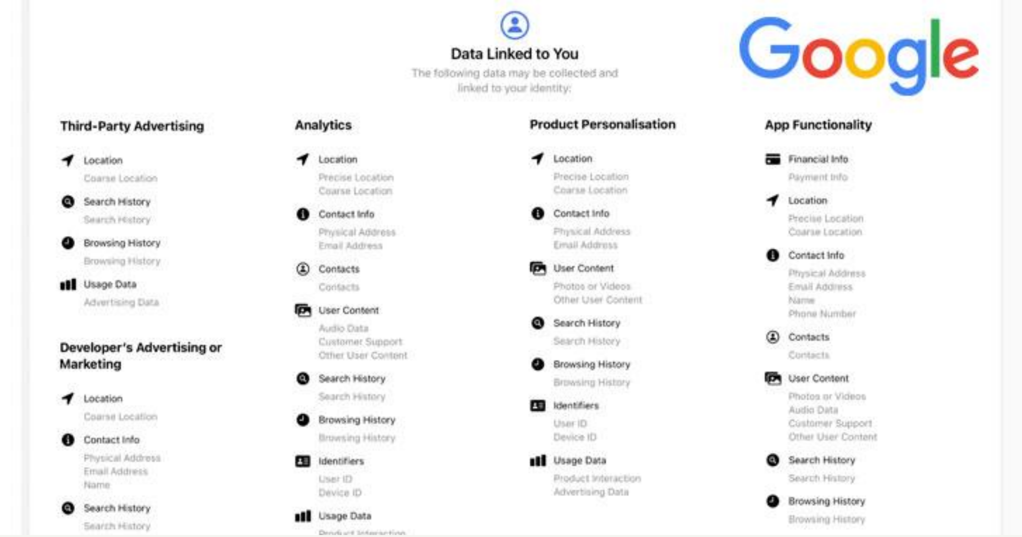
To provide services like spam filtering and live captions, we process your content. For live captions, audio data is temporarily sent to a Google transcription server, but is not linked to any user identifiers or permanently stored.

Privacy in the world of Big Data



Meet uses data to improve your experience

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



Data Linked to You

The following data may be collected and linked to your identity:

Third-Party Advertising

- Location
  Coarse Location
- Search History
  Search History
- Browsing History
  Browsing History
- Usage Data
  Advertising Data

Developer's Advertising or Marketing

- Location
  Coarse Location
- Contact Info
  Physical Address
  Email Address
  Name
- Search History
  Search History

Analytics

- Location
  Precise Location
  Coarse Location
- Contact Info
  Physical Address
  Email Address
- Contacts
  Contacts
- User Content
  Audio Data
  Customer Support
  Other User Content
- Search History
  Search History
- Browsing History
  Browsing History
- Identifiers
  User ID
  Device ID
- Usage Data
  Product Interaction

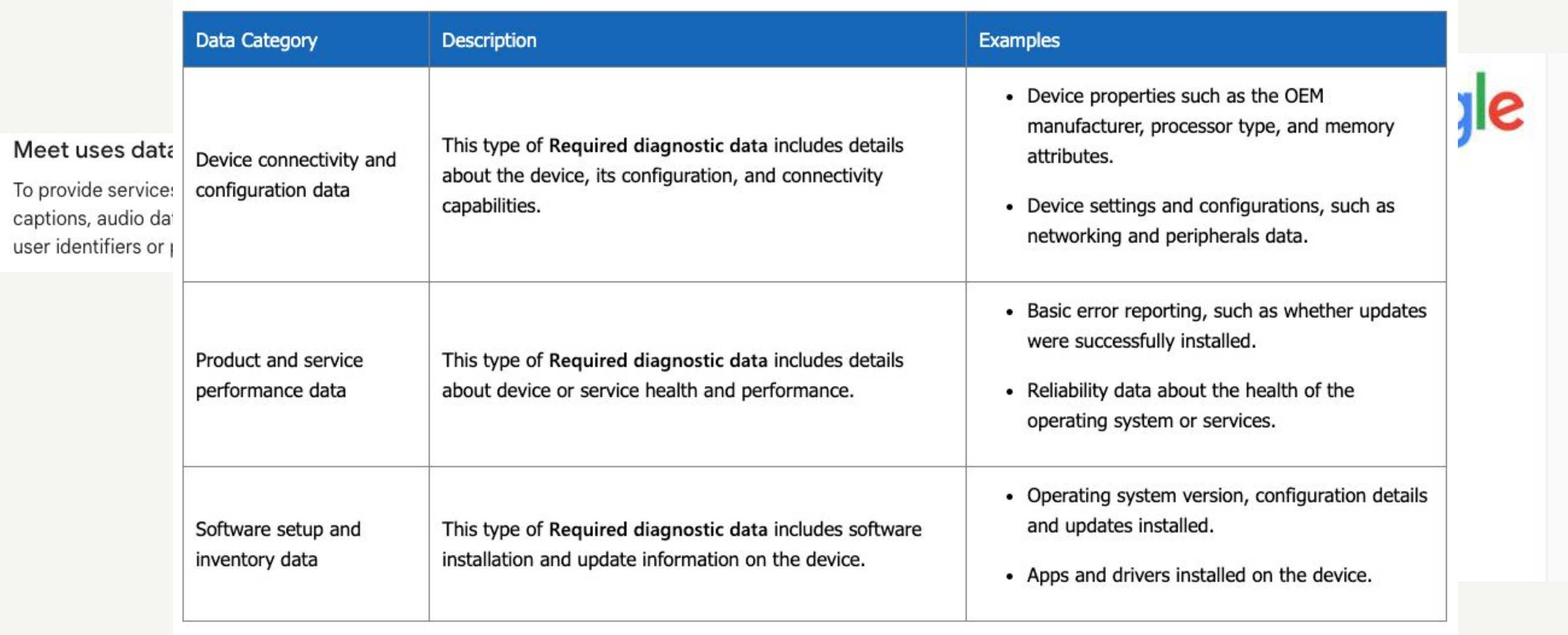
Product Personalisation

- Location
  Precise Location
  Coarse Location
- Contact Info
  Physical Address
  Email Address
- User Content
  Photos or Videos
  Other User Content
- Search History
  Search History
- Browsing History
  Browsing History
- Identifiers
  User ID
  Device ID
- Usage Data
  Product Interaction
  Advertising Data

App Functionality

- Financial Info
  Payment Info
- Location
  Precise Location
  Coarse Location
- Contact Info
  Physical Address
  Email Address
  Name
  Phone Number
- Contacts
  Contacts
- User Content
  Photos or Videos
  Audio Data
  Customer Support
  Other User Content
- Search History
  Search History
- Browsing History
  Browsing History

Privacy in the world of Big Data

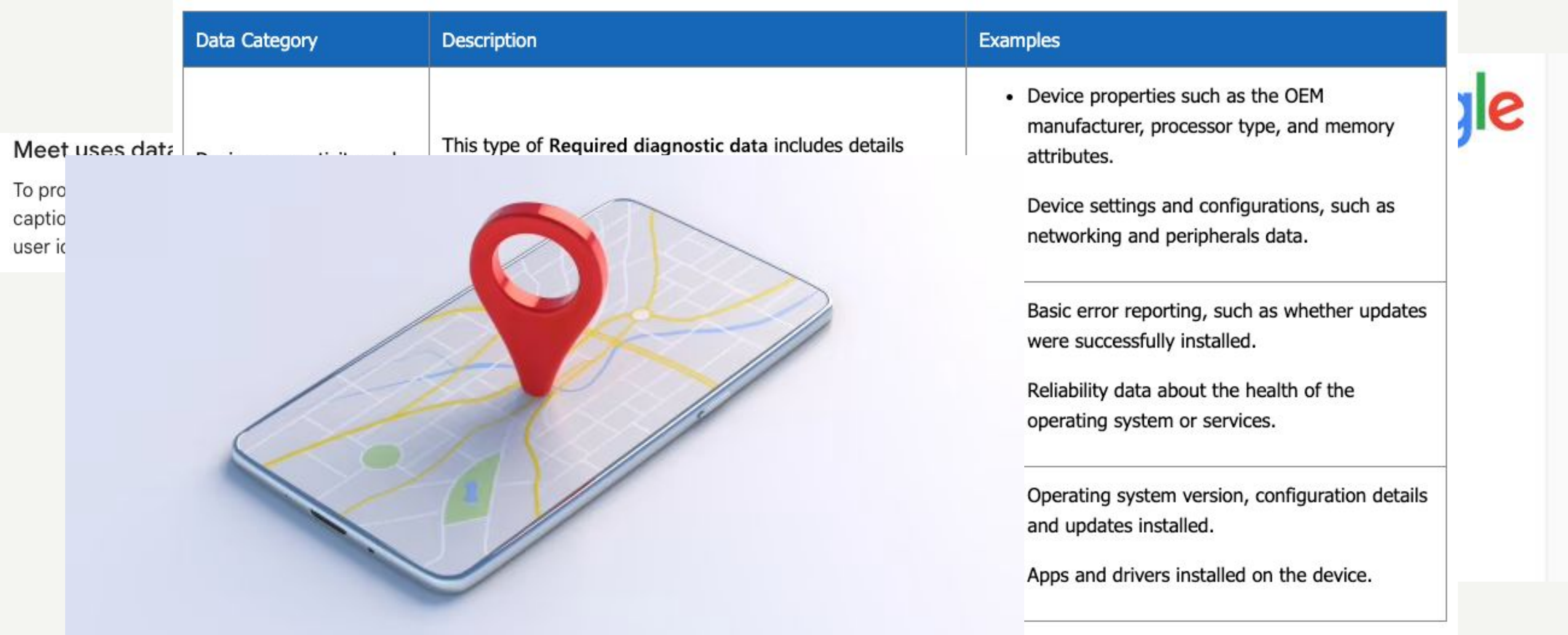


Meet uses data

To provide services
captions, audio dat
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

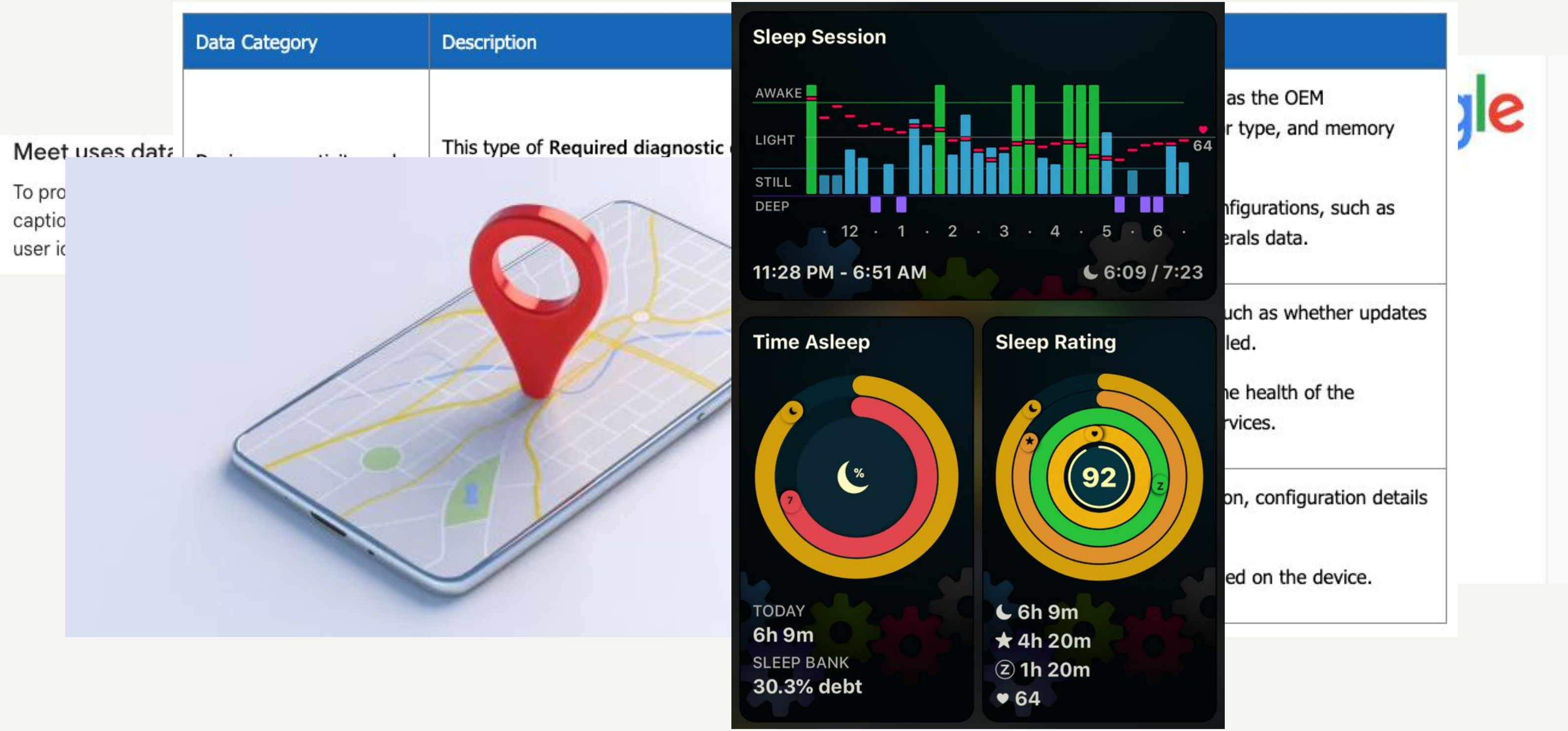


Meet uses data

To pro
captio
user ic

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.

Privacy in the world of Big Data



Meet uses data

To pro
captio
user id

Data Category	Description
D	This type of Required diagnostic

as the OEM
r type, and memory

nfigurations, such as
erals data.

uch as whether updates
led.

e health of the
rvices.

on, configuration details

ed on the device.

gle

Sleep Session

11:28 PM - 6:51 AM

6:09 / 7:23

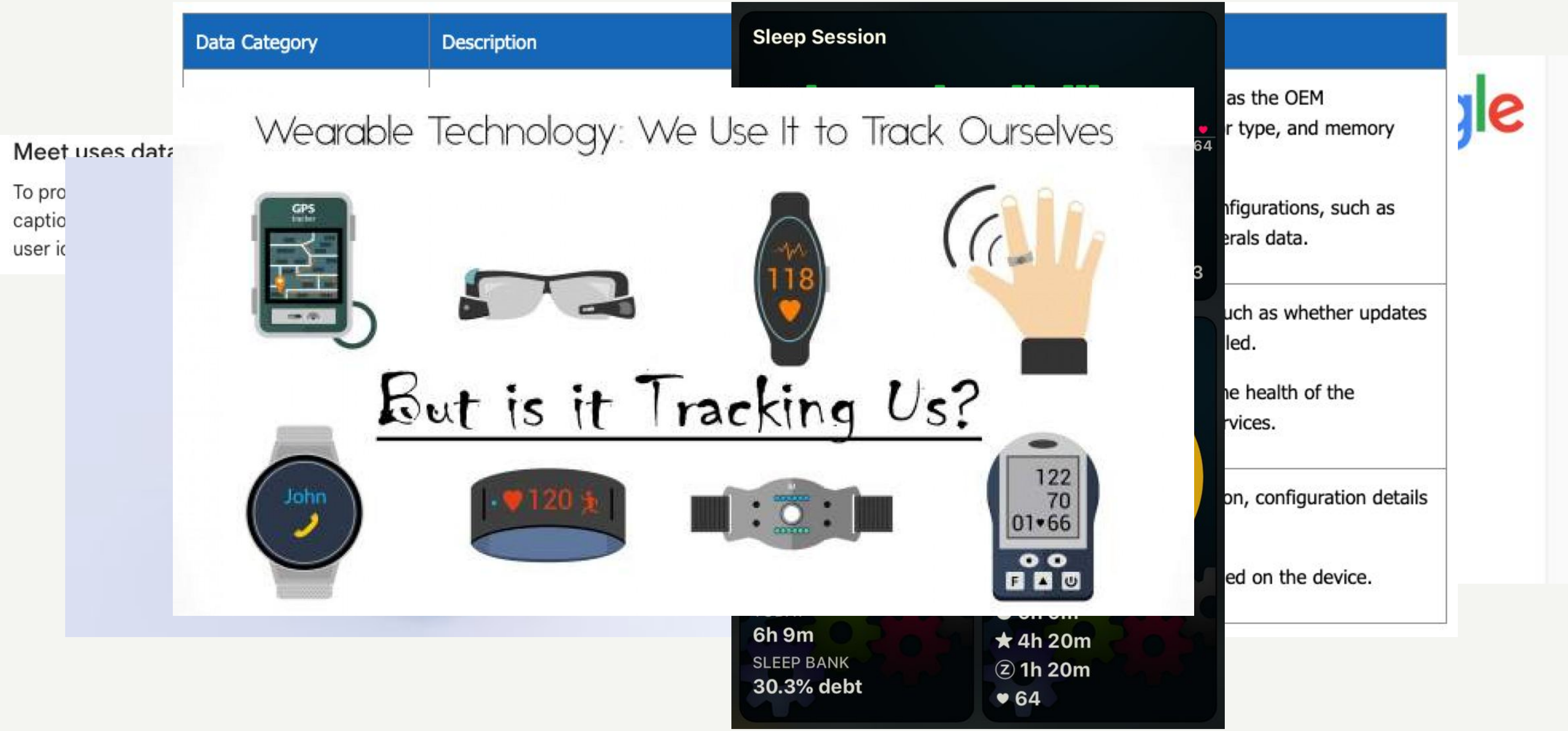
Time Asleep

TODAY
6h 9m
SLEEP BANK
30.3% debt

Sleep Rating

6h 9m
★ 4h 20m
ⓩ 1h 20m
♥ 64

Privacy in the world of Big Data



Data Category | Description

Meet uses data

To pro
captio
user ic

Sleep Session

6h 9m
SLEEP BANK
30.3% debt

★ 4h 20m
ⓩ 1h 20m
♥ 64

as the OEM
r type, and memory

nfigurations, such as
erals data.

uch as whether updates
led.

ne health of the
rvices.

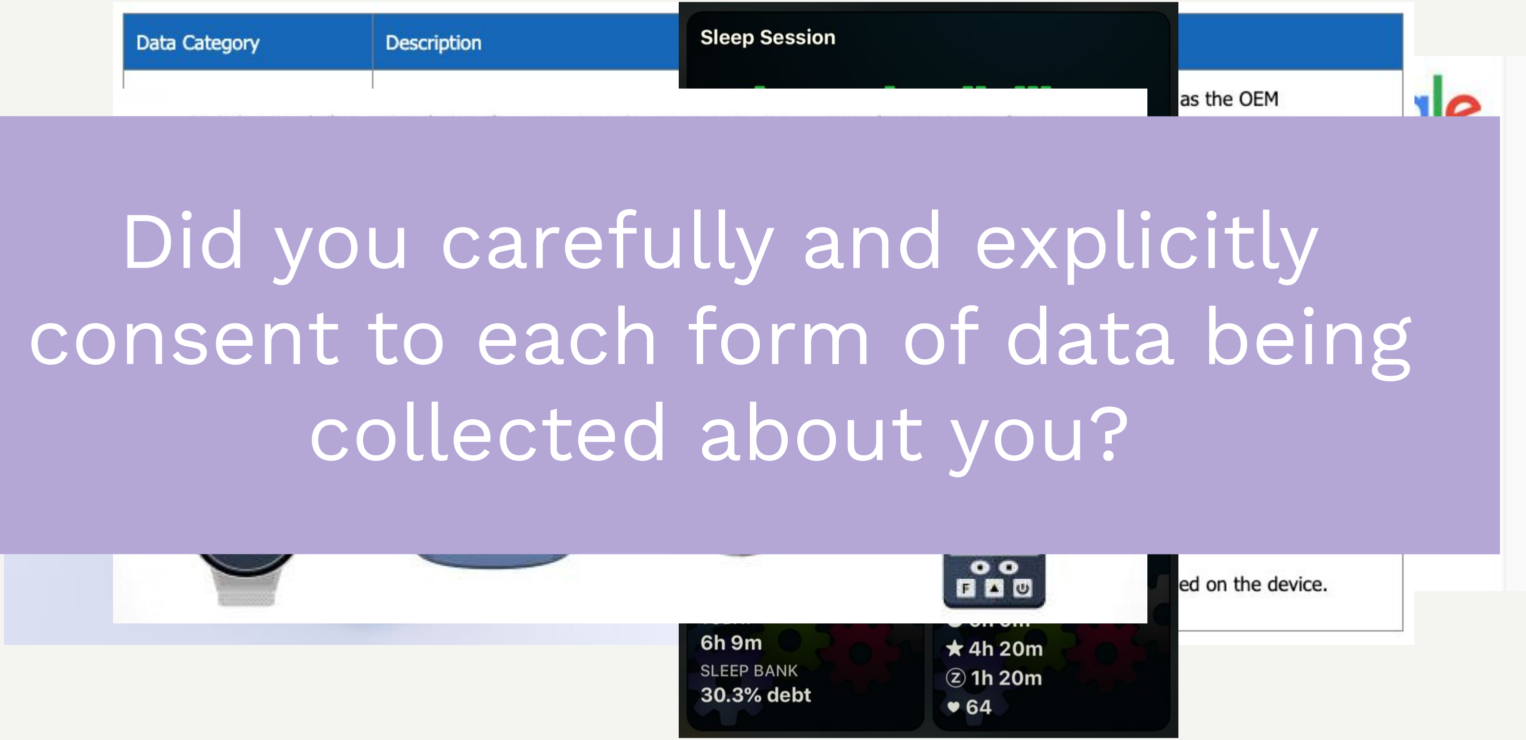
on, configuration details

ed on the device.

Wearable Technology: We Use It to Track Ourselves

But is it Tracking Us?

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

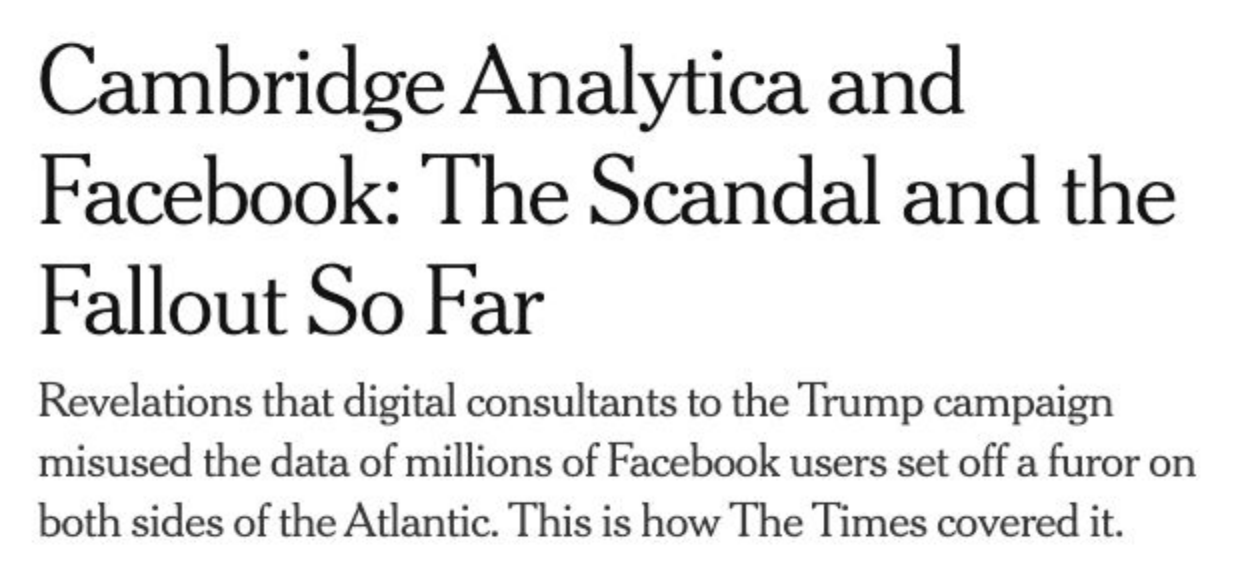
Data Privacy: “*relationship between the collection and dissemination of data, technology, the public expectation of privacy, contextual information norms, and the legal and political issues surrounding them*”

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?

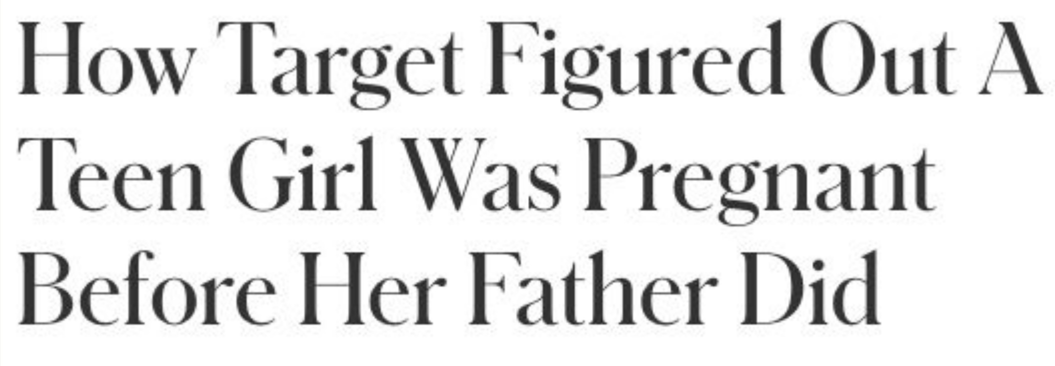
Privacy in the world of Big Data



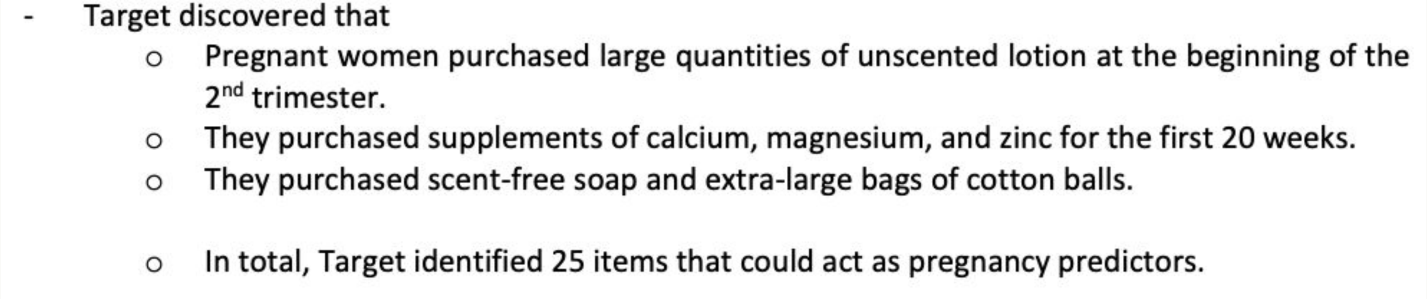
Cambridge Analytica and Facebook: The Scandal and the Fallout So Far

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



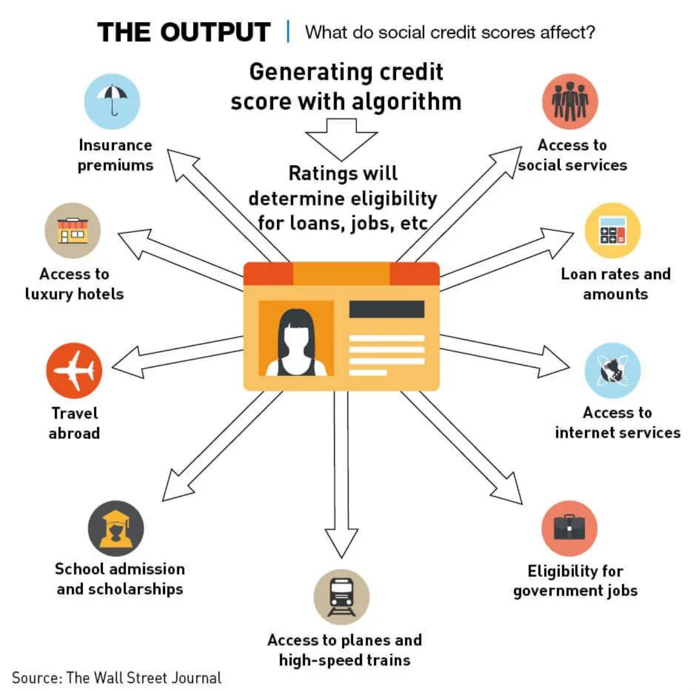
How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



- 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



THE OUTPUT | What do social credit scores affect?

Generating credit score with algorithm

Ratings will determine eligibility for loans, jobs, etc

Insurance premiums

Access to social services

Access to luxury hotels

Loan rates and amounts

Travel abroad

Access to internet services

School admission and scholarships

Access to planes and high-speed trains

Eligibility for government jobs

Source: The Wall Street Journal

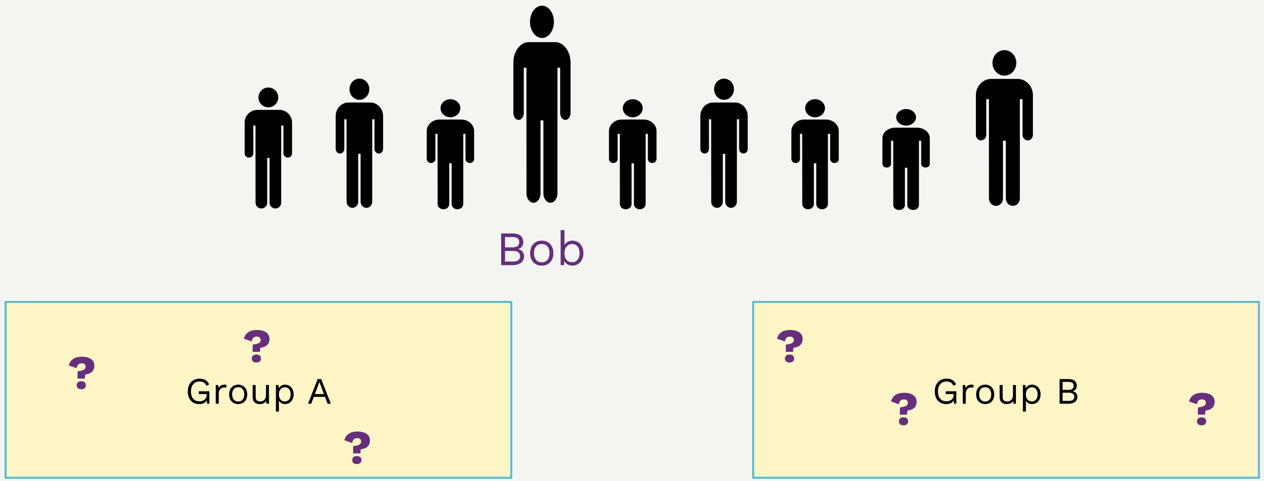
Privacy in the world of Big Data

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

Privacy and Information Flow

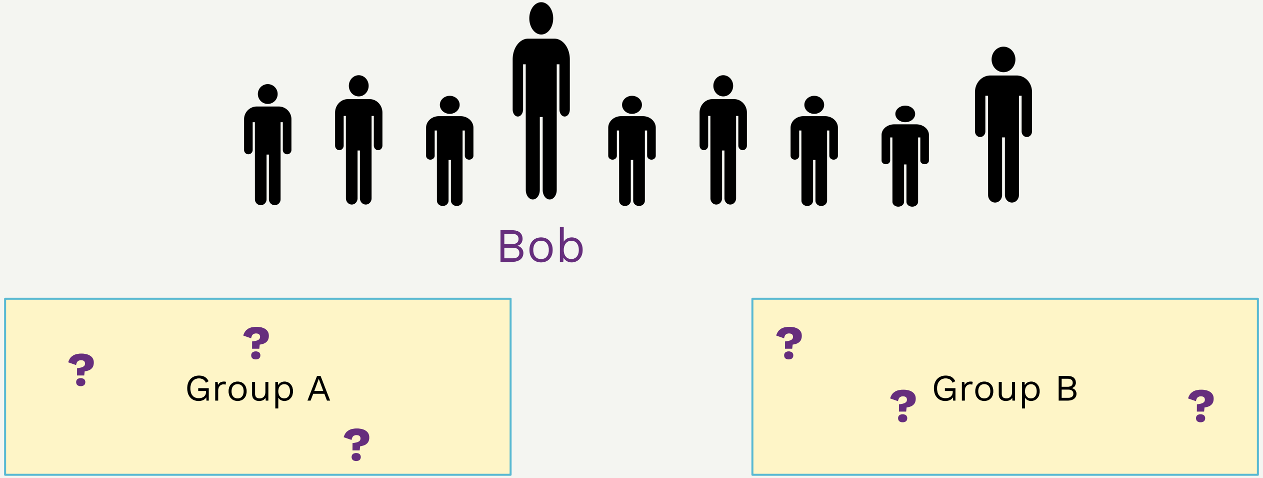
Privacy as Information Leakage

Consider the following example



Privacy as Information Leakage

Consider the following example

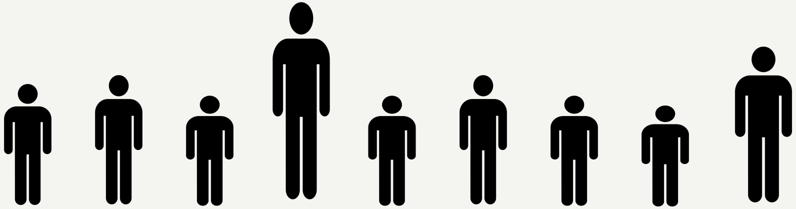


Average Height = 5’10

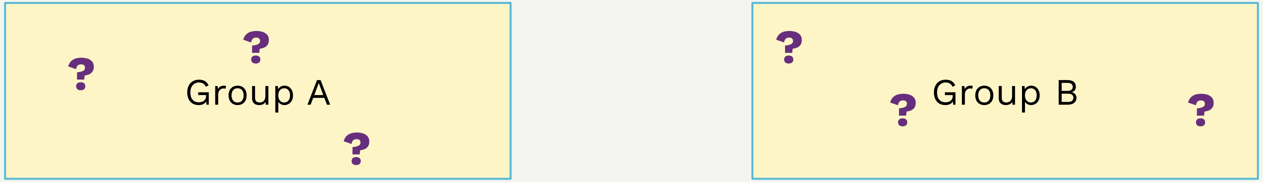
Average Height = 5’6

Privacy as Information Leakage

Consider the following example



Bob

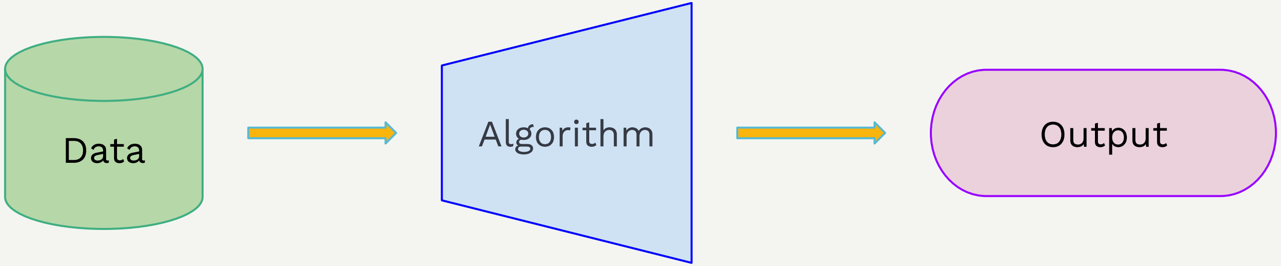


Average Height = 5’10

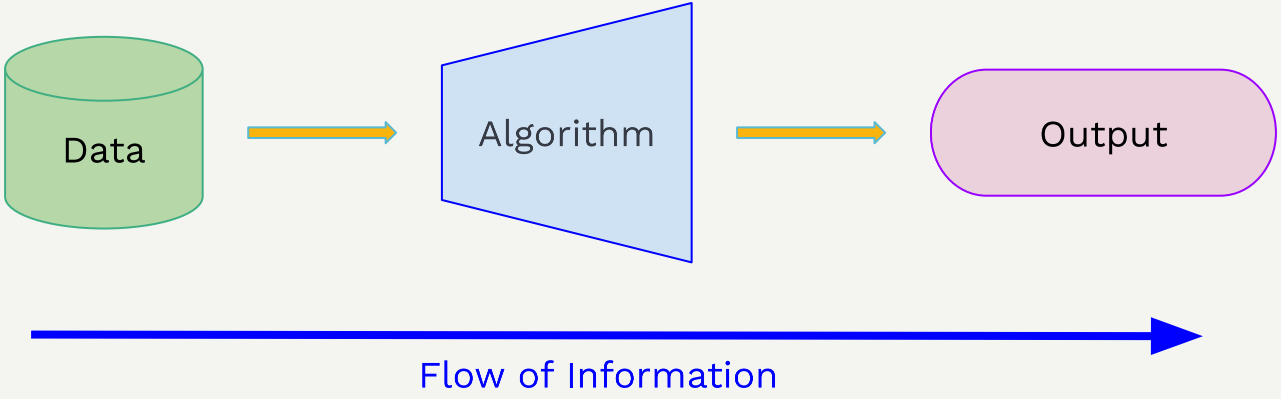
Average Height = 5’6

Can you guess which group Bob belongs to?

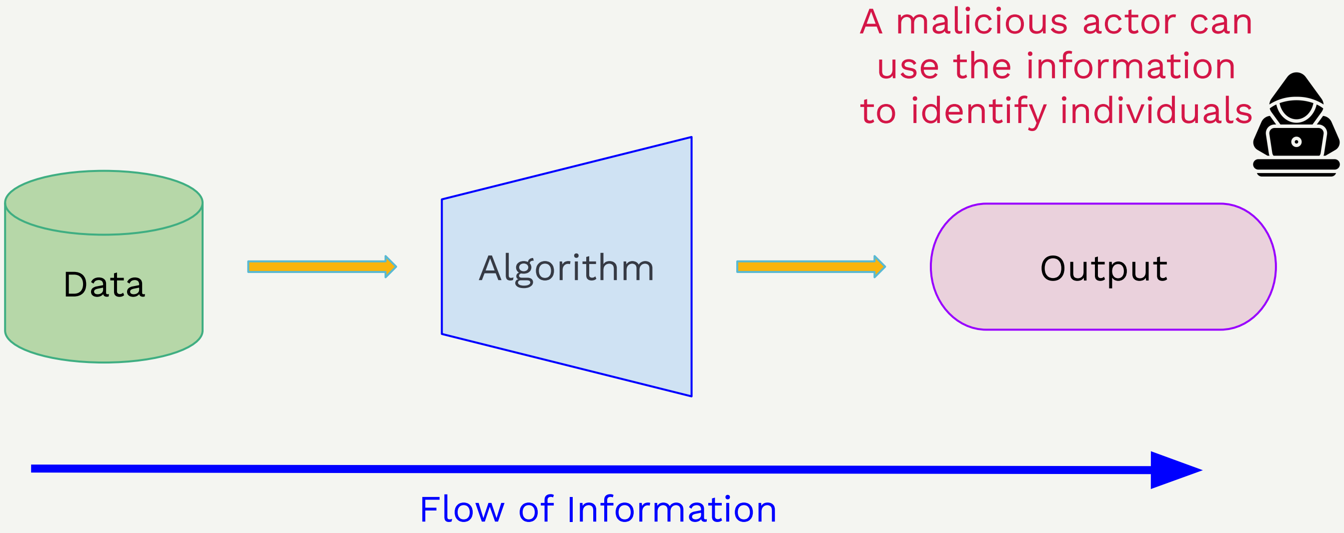
Privacy as Information Leakage



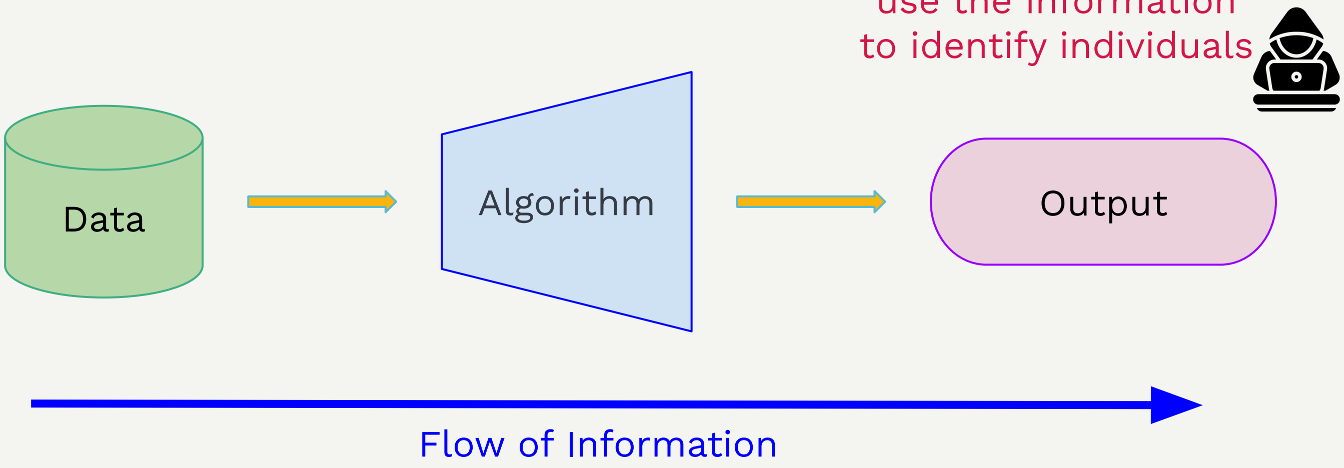
Privacy as Information Leakage



Privacy as Information Leakage


A malicious actor can use the information to identify individuals
Data
Algorithm
Output
Flow of Information

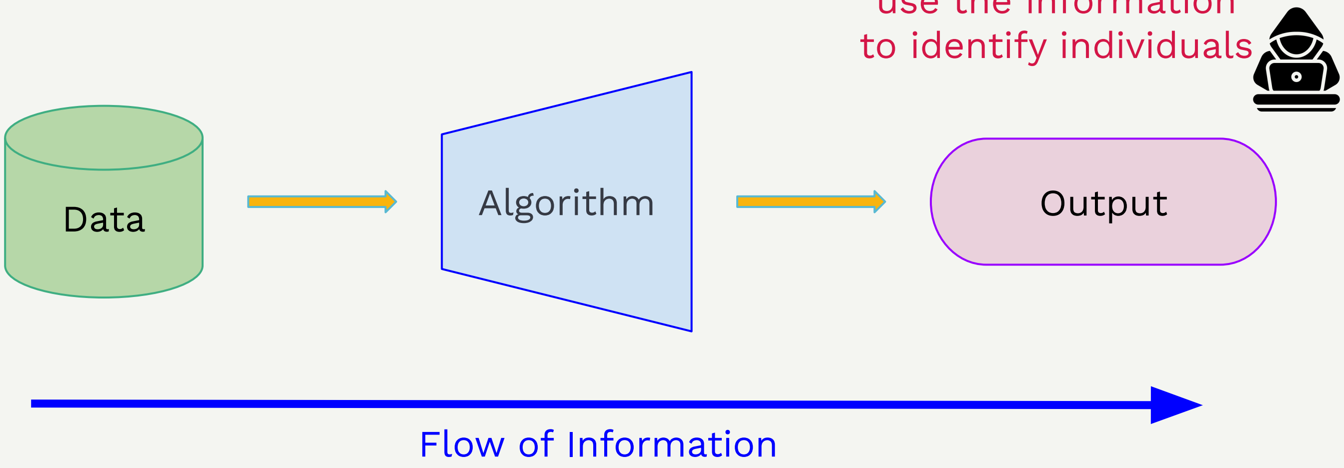
Privacy as Information Leakage



A malicious actor can use the information to identify individuals

What if there is a perfect flow of information from the data to the output?

Privacy as Information Leakage

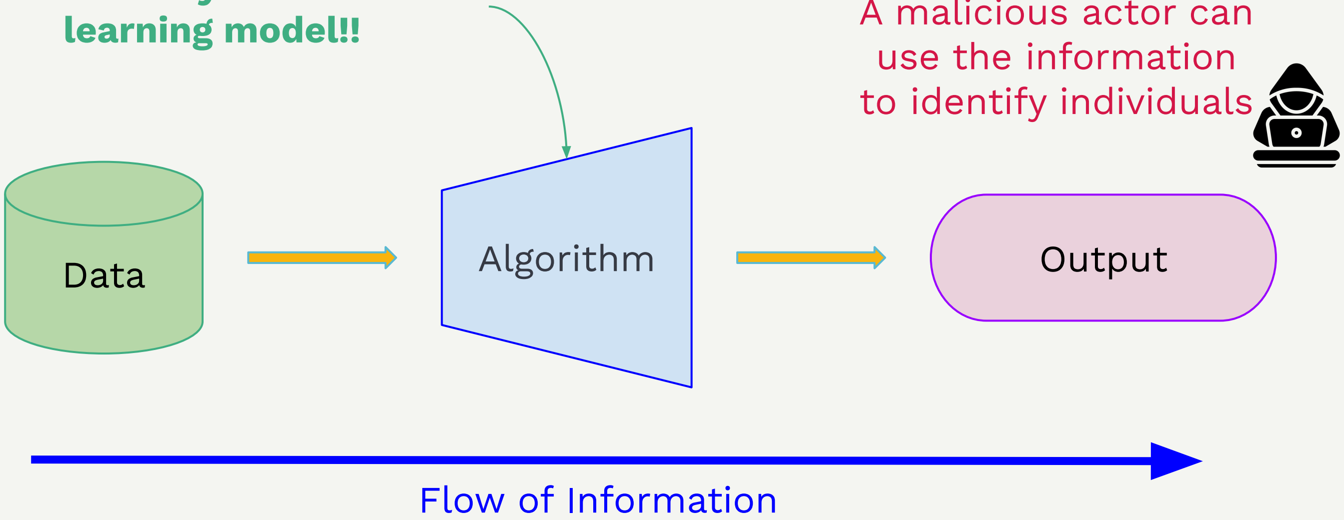


A malicious actor can use the information to identify individuals

What if there is a perfect flow of information from the data to the output?

What if there is NO flow of information from the data to the output?

Privacy as Information Leakage


This can be your machine learning model!!
A malicious actor can use the information to identify individuals
Data
Algorithm
Output
Flow of Information

What if there is a perfect flow of information from the data to the output?

What if there is NO flow of information from the data to the output?

Privacy and Security

Data Privacy: *rights of individuals to control how their personal information is collected, used, and shared.*

Data Security: *protecting data from external and internal threats to ensure its integrity, confidentiality, and availability.*

Information Leakage through ML Models

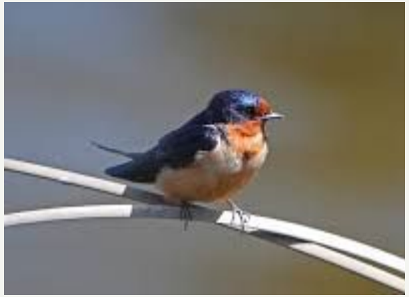
Generalization and Memorization

Generalization and Memorization

barn swallow

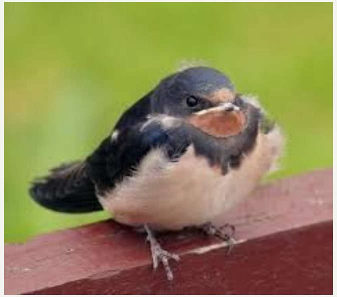
tree swallow

Generalization and Memorization

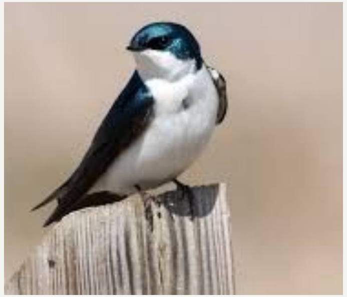


Can you tell me if this is a barn swallow or a tree swallow?

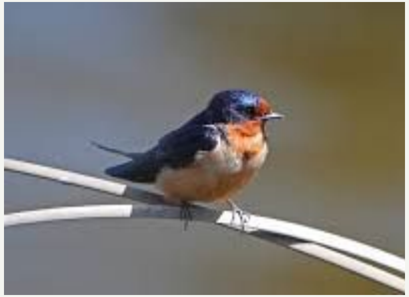
Generalization and Memorization



barn swallow



tree swallow



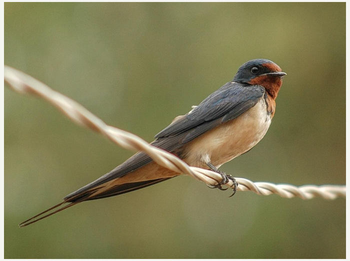
Can you tell me if this is a barn swallow or a tree swallow?

Generalization and Memorization

barn swallow

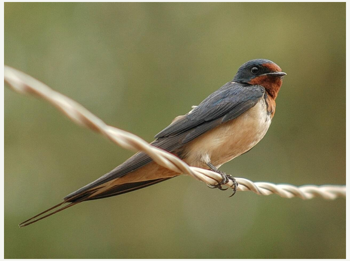
tree swallow

Generalization and Memorization

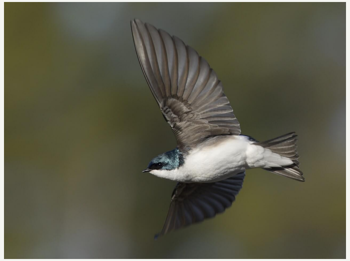


Can you tell me if this is a barn swallow or a tree swallow?

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.*

Generalization and Memorization

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

Given enough data, generalization can rely on extracting ‘patterns’ from the data. **But this isn’t always possible!**

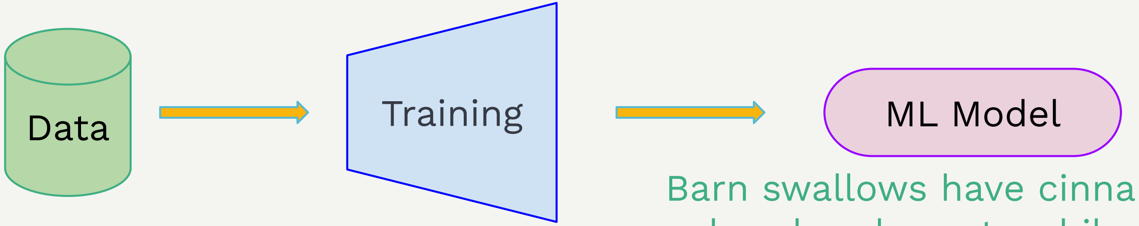
Generalization and Memorization

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

Given enough data, generalization can rely on extracting ‘patterns’ from the data. **But this isn’t always possible!**

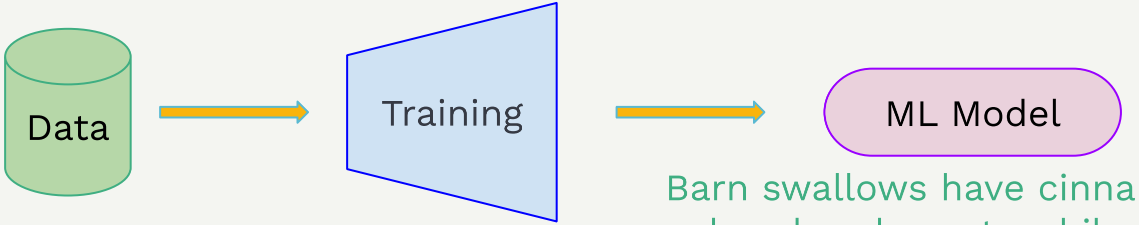
Thus, machine learning models need to memorize.

Information Leakage through ML Models



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

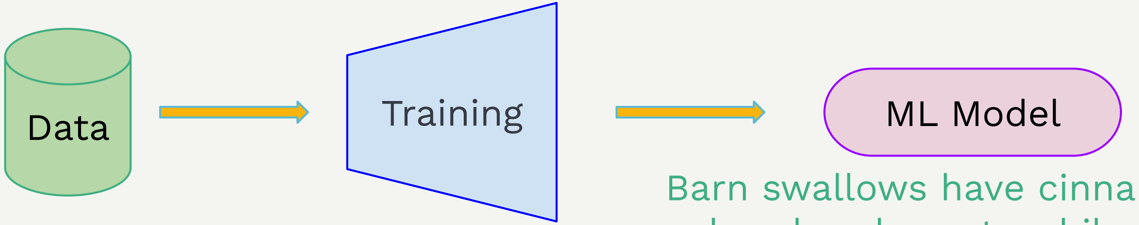
Information Leakage through ML Models



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

Was this image part of your data?

Information Leakage through ML Models

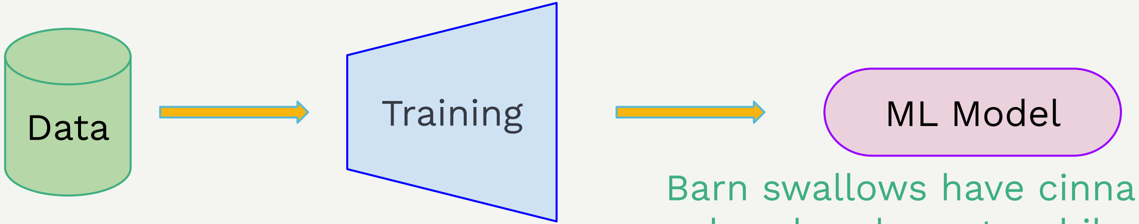


Barn swallows have cinnamon colored underparts while tree swallows have white underparts.

Was this image part of your data?

Can’t tell

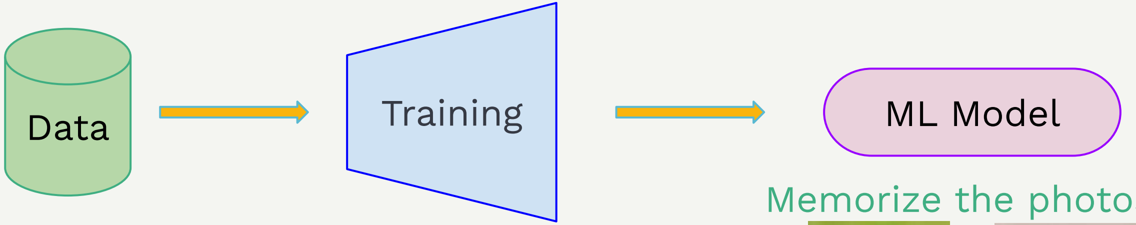
Information Leakage through ML Models



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

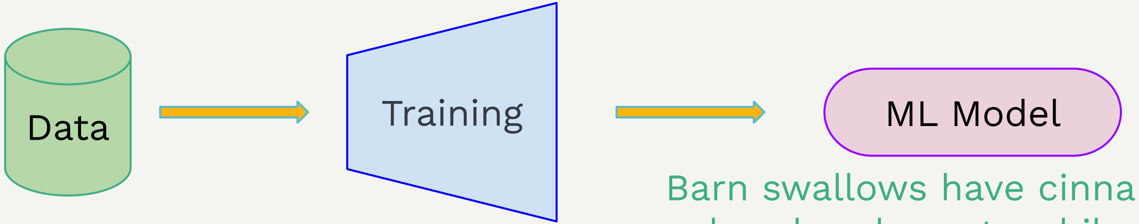
Was this image part of your data?

Can’t tell



Memorize the photos!

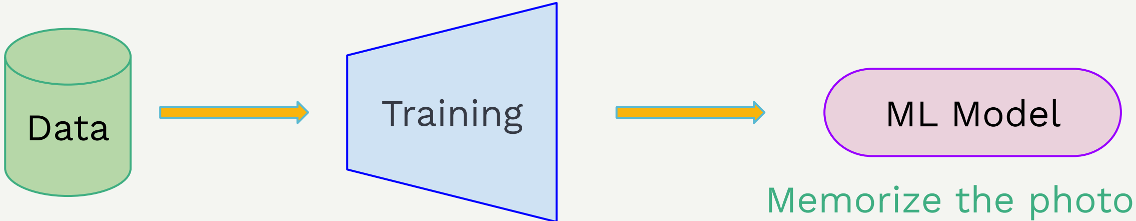
Information Leakage through ML Models



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

Was this image part of your data?

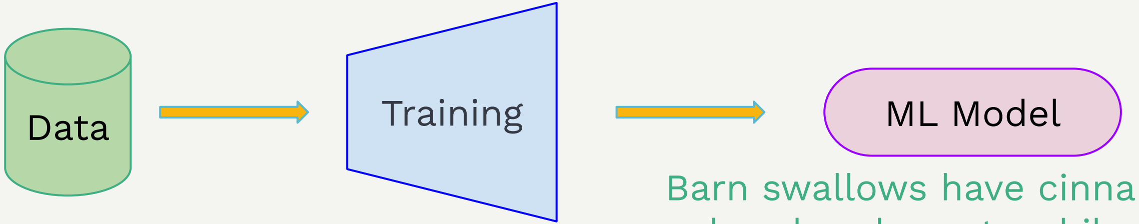
Can’t tell



Was this image part of your data?

Memorize the photos!

Information Leakage through ML Models



Data

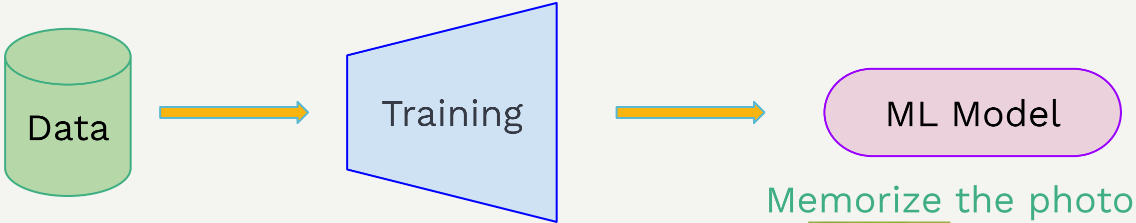
Training

ML Model

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

Was this image part of your data?

Can’t tell



Data

Training

ML Model

Memorize the photos!

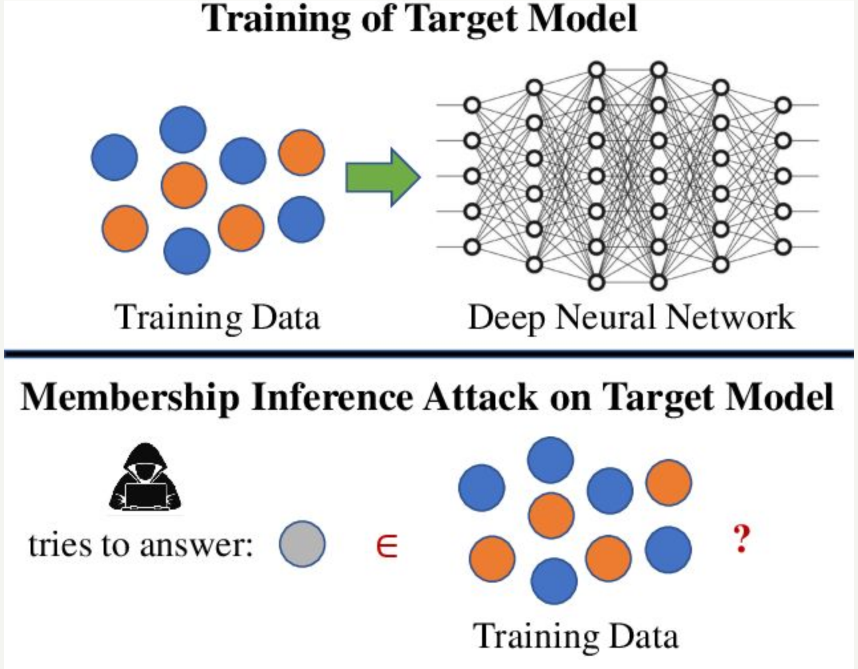
Was this image part of your data?

Yes it was

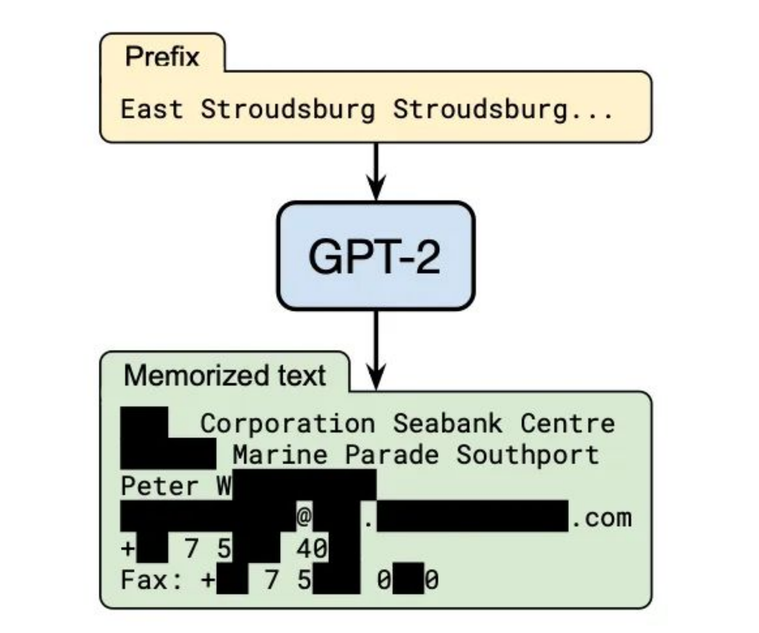
Information Leakage through ML Models

Memorization in ML models is sometimes necessary to perform well, but can lead to information leakage and privacy risks!

Membership Inference Attacks



Data Extraction Attacks



Prefix

East Stroudsburg Stroudsburg...

GPT-2

Memorized text

[redacted] Corporation Seabank Centre
Marine Parade Southport
Peter W[redacted]
[redacted]@[redacted].[redacted].com
+[redacted] 7 5[redacted] 40
Fax: +[redacted] 7 5[redacted] 0[redacted]0

Model Extraction Attacks

De-Anonymization Attacks

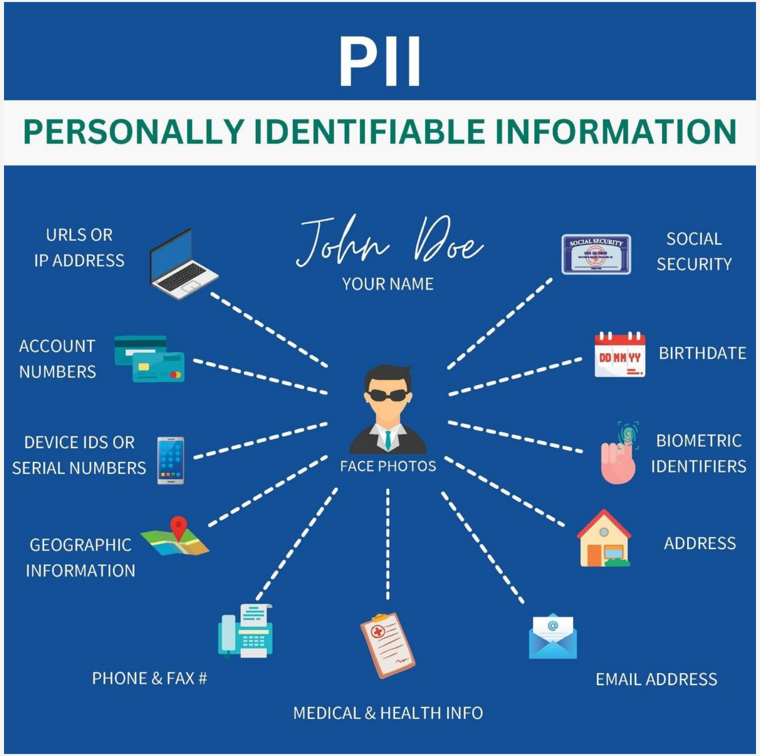
Property Inference Attacks

Side-Channel Attacks

.....

Technical Frameworks in Data Privacy

Personally Identifiable Information (PII)



The Promise of Anonymization

Name	Zipcode	Age	Gender	Genetic Marker for Cancer
Alice	117068	27	F	Yes
Bob	167056	64	M	No
Charlie	118567	32	M	No
David	191504	81	M	No

Insurance companies: Who has genetic markers for cancer? I would like to raise their premium and get more money!

Alice has genetic markers for cancer.

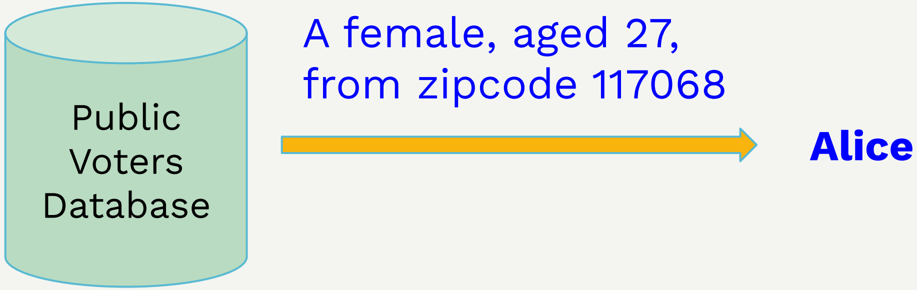
The Promise of Anonymization

Name	Zipcode	Age	Gender	Genetic Marker for Cancer
	117068	27	F	Yes
	167056	64	M	No
	118567	32	M	No
	191504	81	M	No

Insurance companies: Who has genetic markers for cancer? I would like to raise their premium and get more money!

A female, aged 27, from zipcode 117068, has genetic markers for cancer.

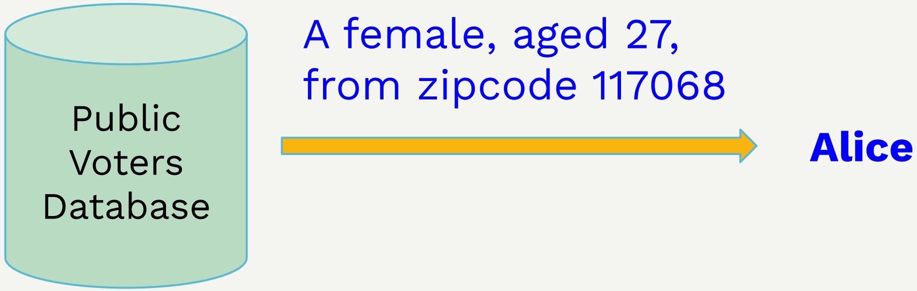
The Promise of Anonymization



Insurance companies: Who has genetic markers for cancer? I would like to raise their premium and get more money!

A female, aged 27, from zipcode 117068, has genetic markers for cancer.

The Promise of Anonymization



Public Voters Database

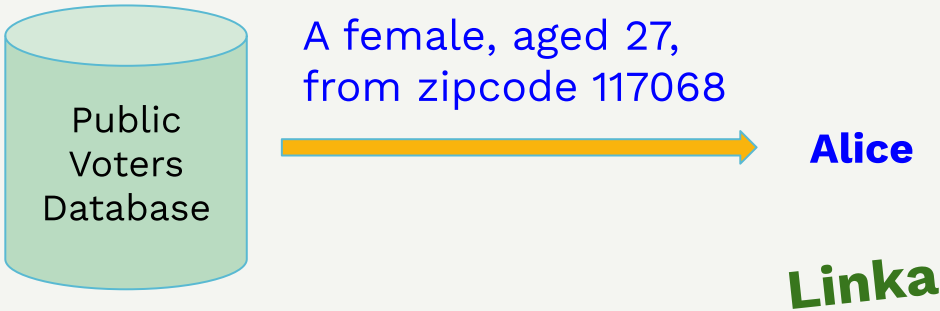
A female, aged 27, from zipcode 117068

Alice

Insurance companies: Who has genetic markers for cancer? I would like to raise their premium and get more money!

Alice has genetic markers for cancer.

The Promise of Anonymization



Public Voters Database

A female, aged 27, from zipcode 117068

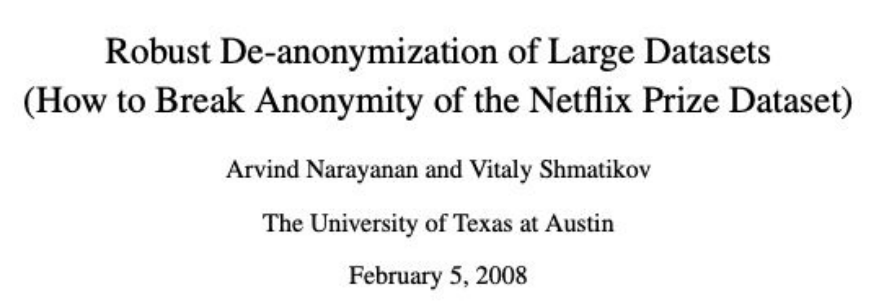
Alice

Linkage Attacks

Insurance companies: Who has genetic markers for cancer? I would like to raise their premium and get more money!

Alice has genetic markers for cancer.

The Promise of Anonymization

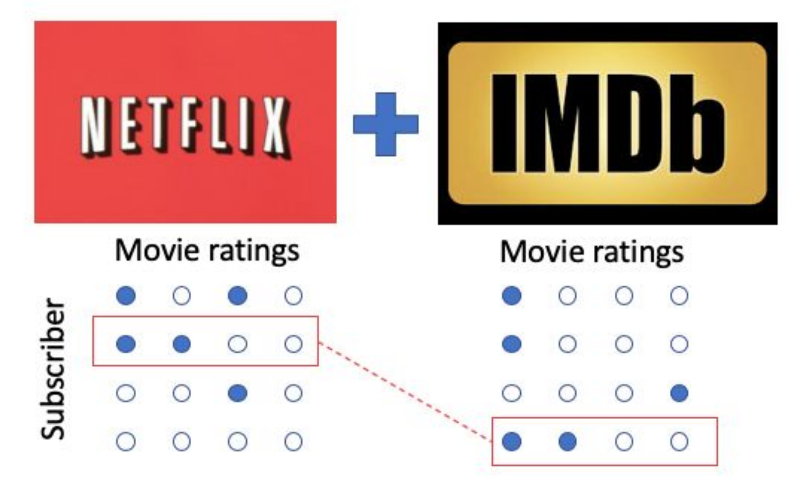


Robust De-anonymization of Large Datasets
(How to Break Anonymity of the Netflix Prize Dataset)

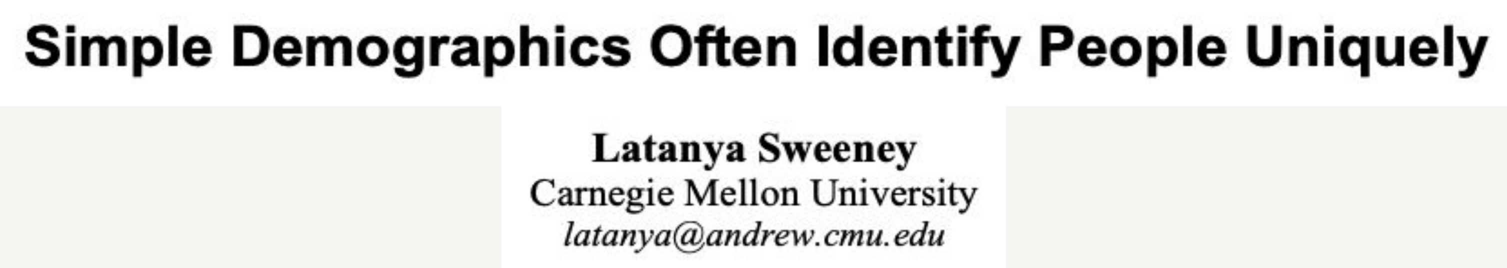
Arvind Narayanan and Vitaly Shmatikov

The University of Texas at Austin

February 5, 2008

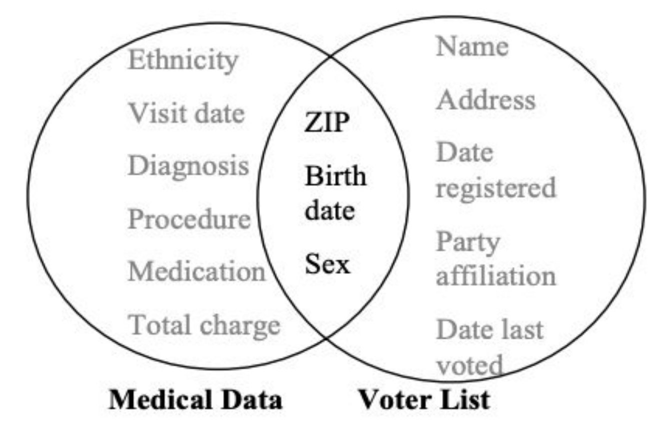


The Promise of Anonymization

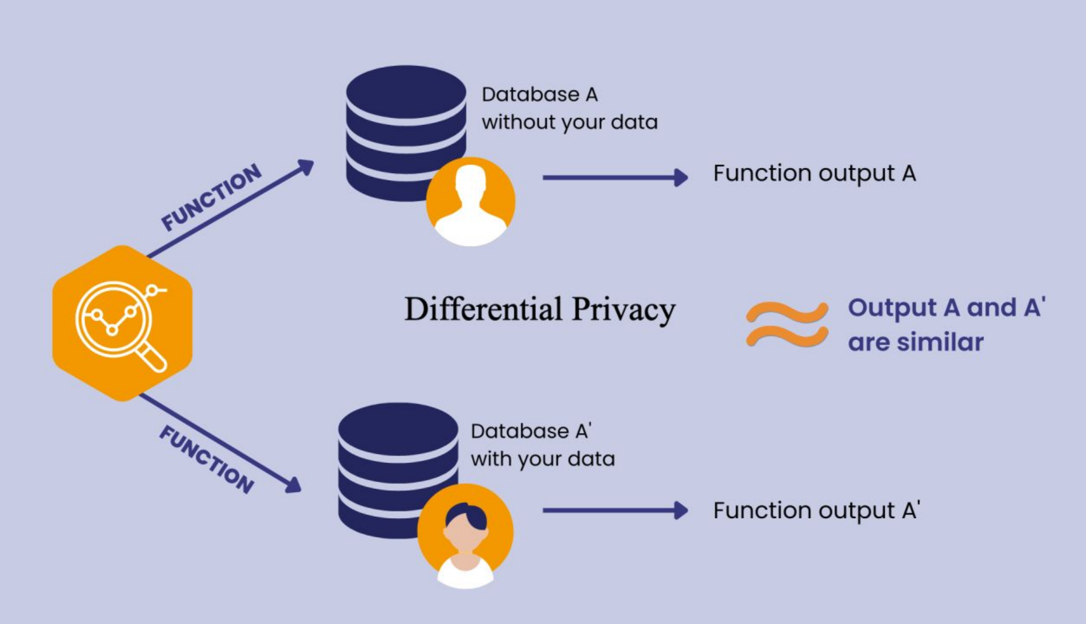


Simple Demographics Often Identify People Uniquely

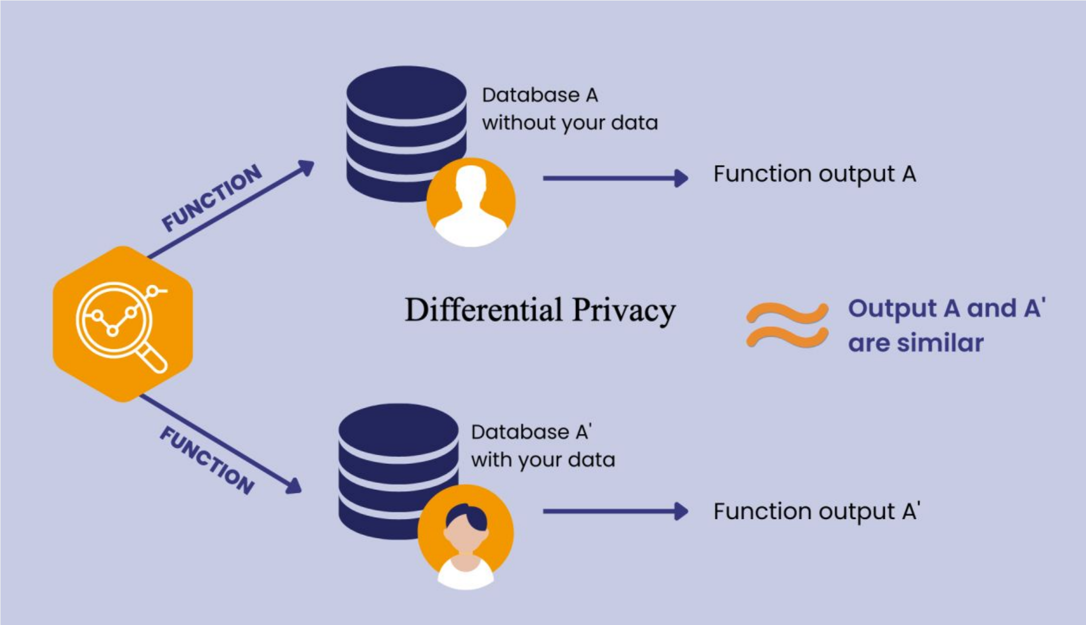
Latanya Sweeney
Carnegie Mellon University
latanya@andrew.cmu.edu



Differential Privacy



Differential Privacy



If A and A' are similar, it is not possible to detect the membership of your data, IRRESPECTIVE of any additional data available!!

Other Concepts in Privacy

Other Concepts in Privacy

Federated Learning: *multiple entities coming together to collaboratively train models while ensuring that their data remains decentralized.*

Other Concepts in Privacy

Federated Learning: *multiple entities coming together to collaboratively train models while ensuring that their data remains decentralized.*

Homomorphic Encryption: performing *complex mathematical operations on encrypted data without compromising the encryption.*

Other Concepts in Privacy

Federated Learning: *multiple entities coming together to collaboratively train models while ensuring that their data remains decentralized.*

Homomorphic Encryption: performing *complex mathematical operations on encrypted data without compromising the encryption.*

Privacy by Design Principles: *proactively embedding privacy in ML systems, to anticipate and prevent privacy invasive events before they occur.*

Other Concepts in Privacy

Federated Learning: *multiple entities coming together to collaboratively train models while ensuring that their data remains decentralized.*

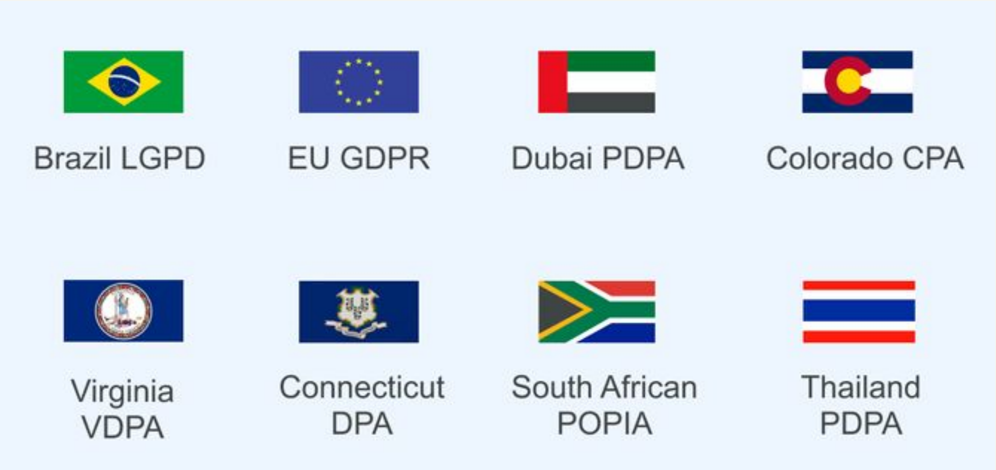
Homomorphic Encryption: performing *complex mathematical operations on encrypted data without compromising the encryption.*

Privacy by Design Principles: *proactively embedding privacy in ML systems, to anticipate and prevent privacy invasive events before they occur.*

Secure Multi-Party Computations, Zero Knowledge Proofs, ...

Legal Frameworks in Data Privacy

Data Protection Regulations



Data Protection Regulations

2020 Brazil LGPD

2018 EU GDPR

2022 Dubai PDPA

2023 Colorado CPA

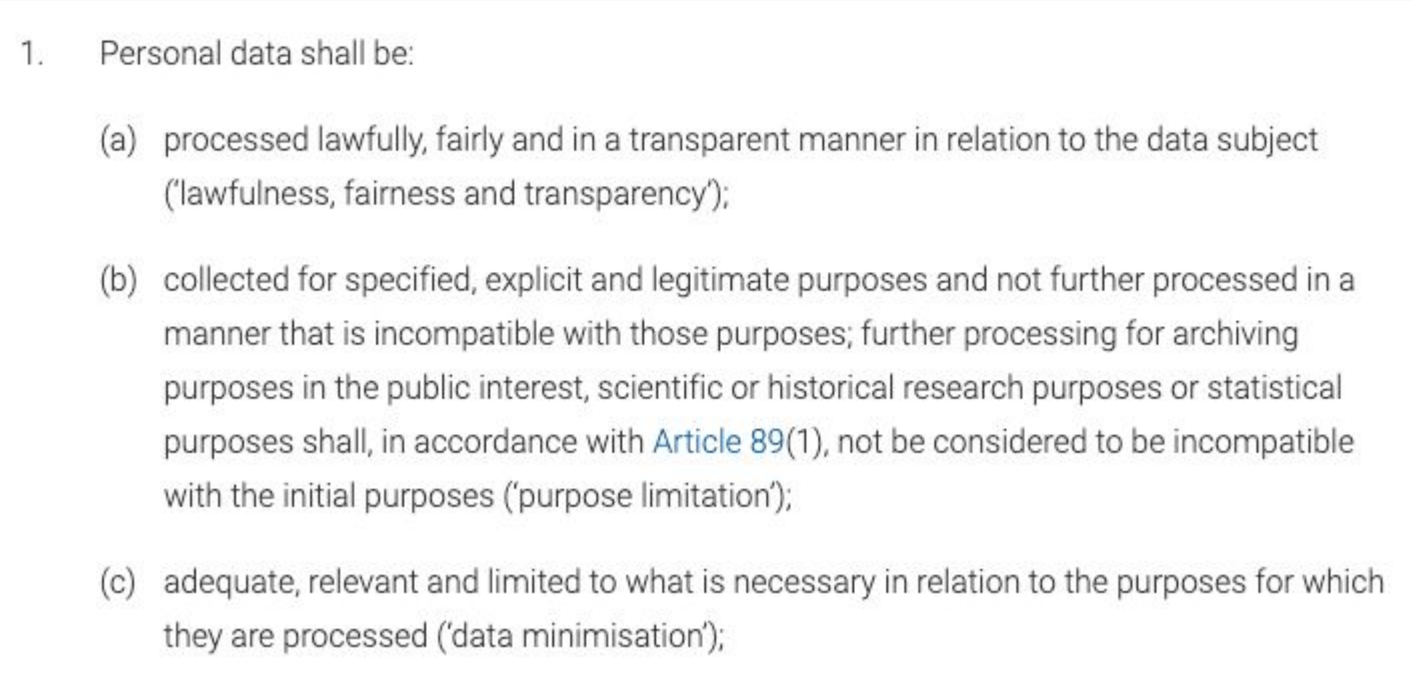
2021 Virginia VDPA

2023 Connecticut DPA

2020 South African POPIA

2022 Thailand PDPA

Data Minimization



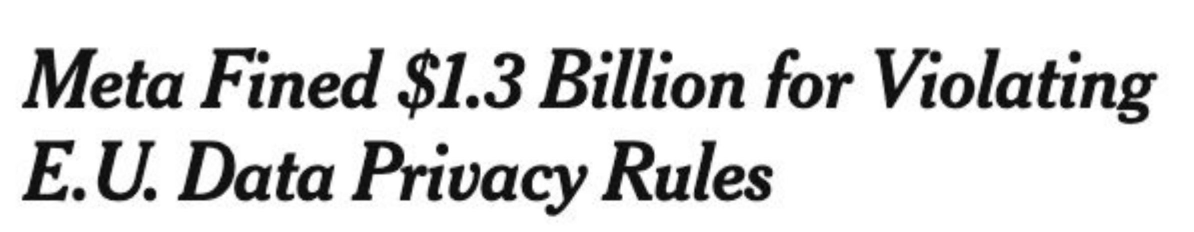
1. Personal data shall be:

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

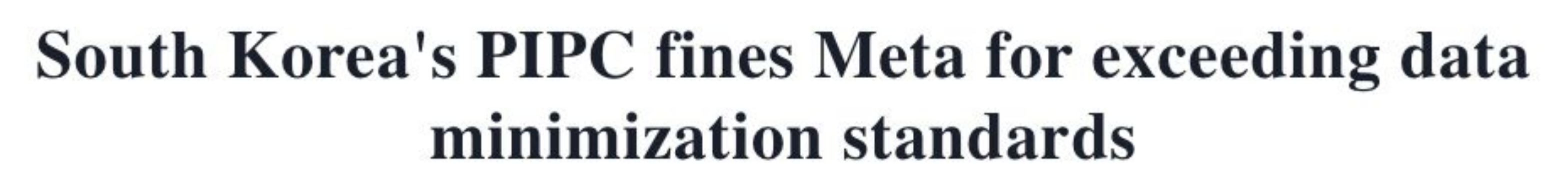
   (b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with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

Data Minimization

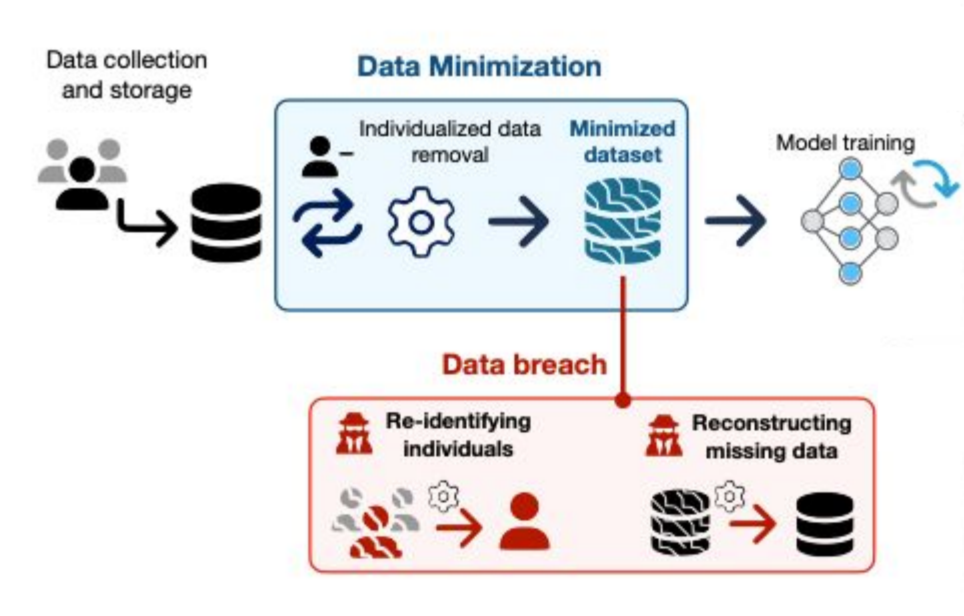


Meta Fined $1.3 Billion for Violating E.U. Data Privacy Rules

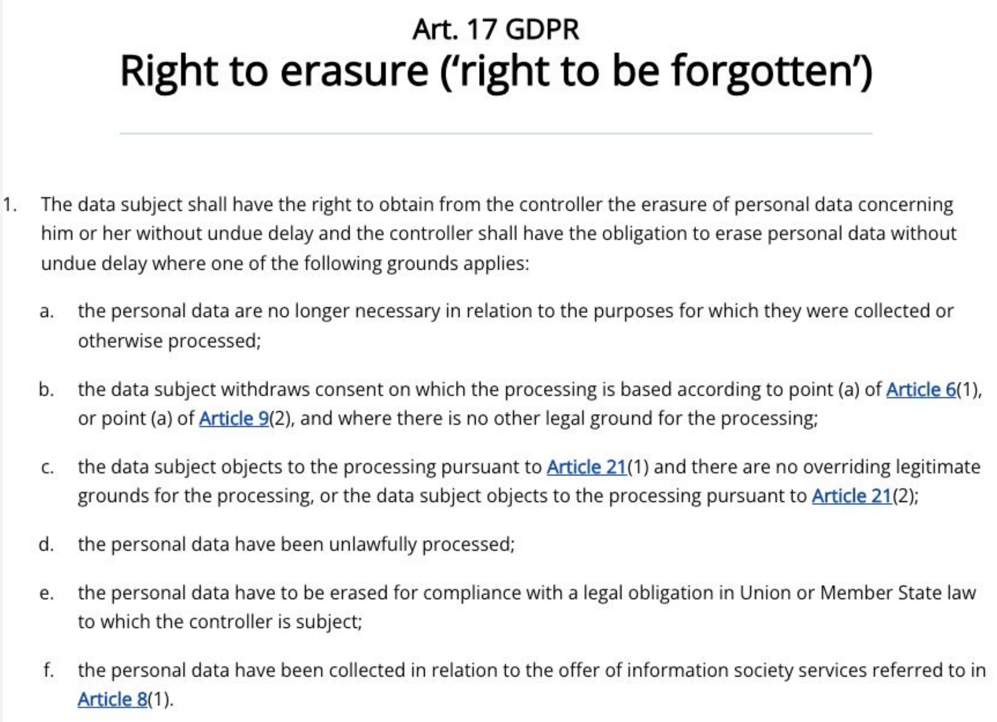


South Korea's PIPC fines Meta for exceeding data minimization standards

Data Minimization


Data collection and storage
Data Minimization
Individualized data removal
Minimized dataset
Model training
Data breach
Re-identifying individuals
Reconstructing missing data

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;

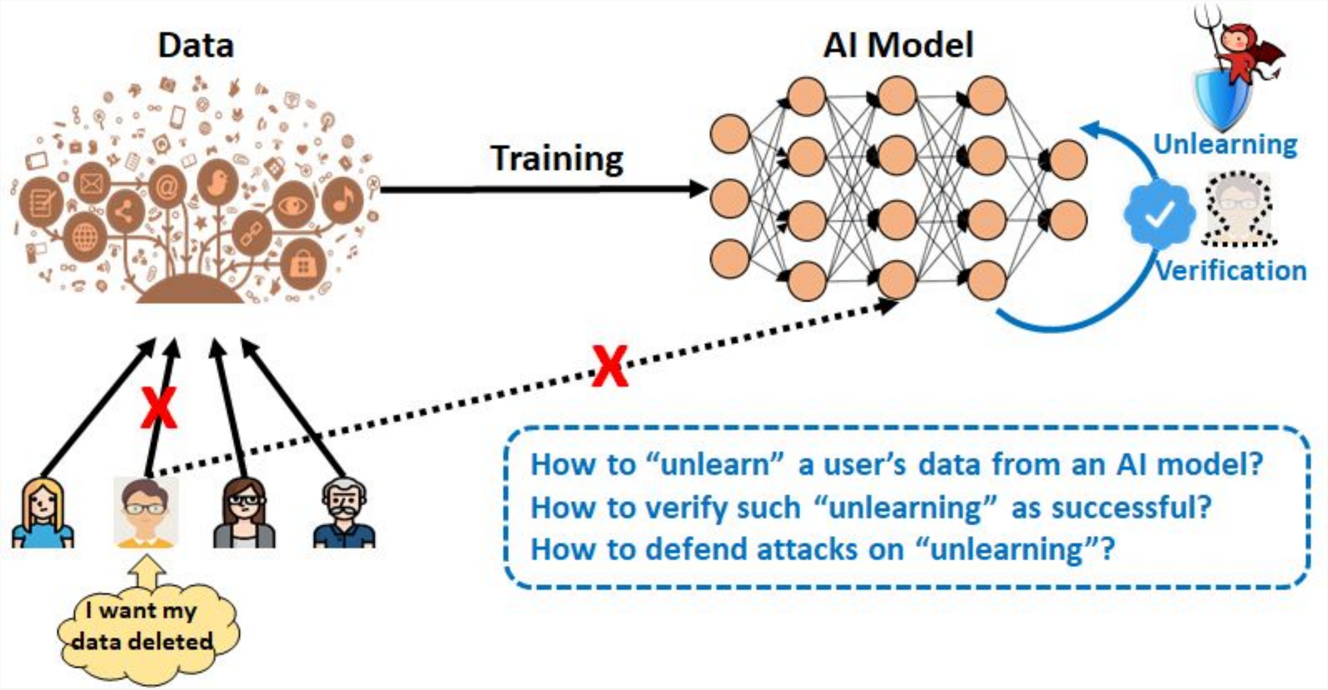
   c. the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);

   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



How to “unlearn” a user’s data from an AI model?
How to verify such “unlearning” as successful?
How to defend attacks on “unlearning”?

Right to be informed

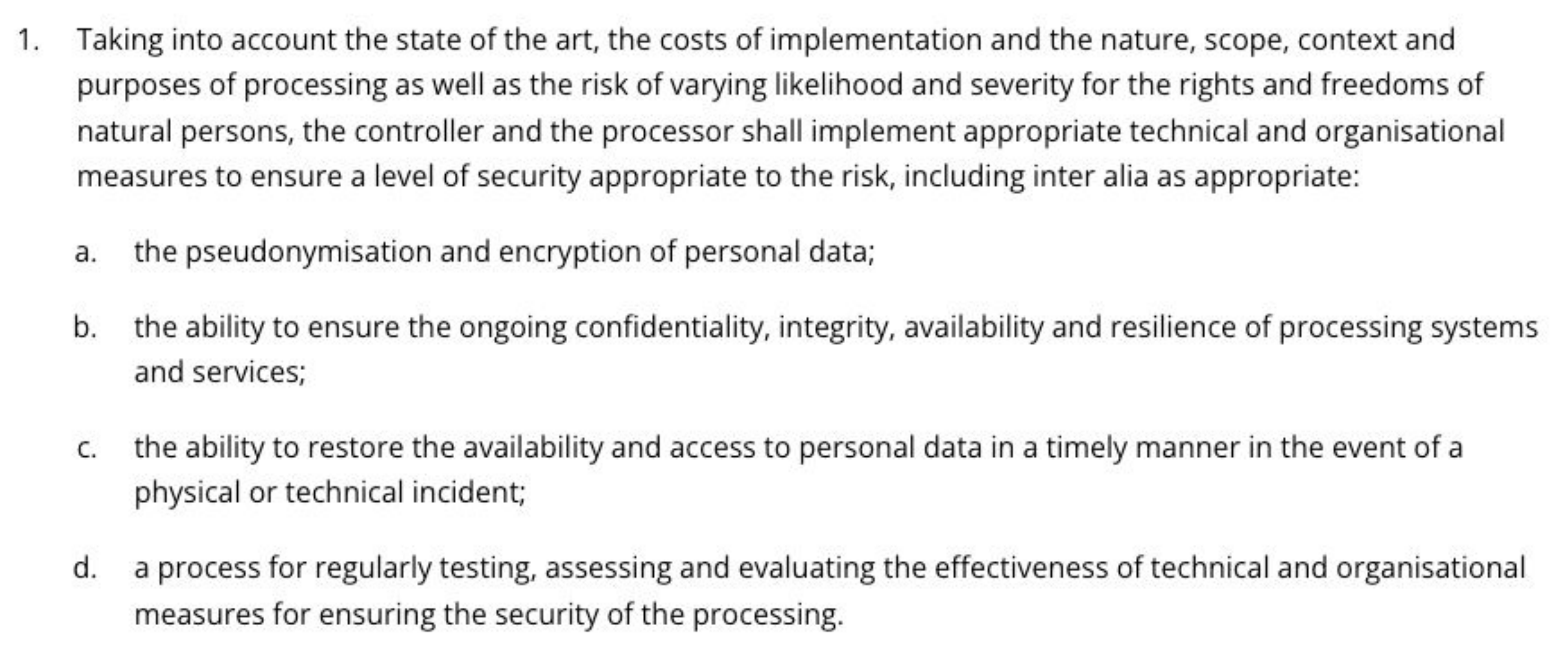
Right of access

Right to restrict processing

Right to rectification

.....

Security Safeguard Requirements



1. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

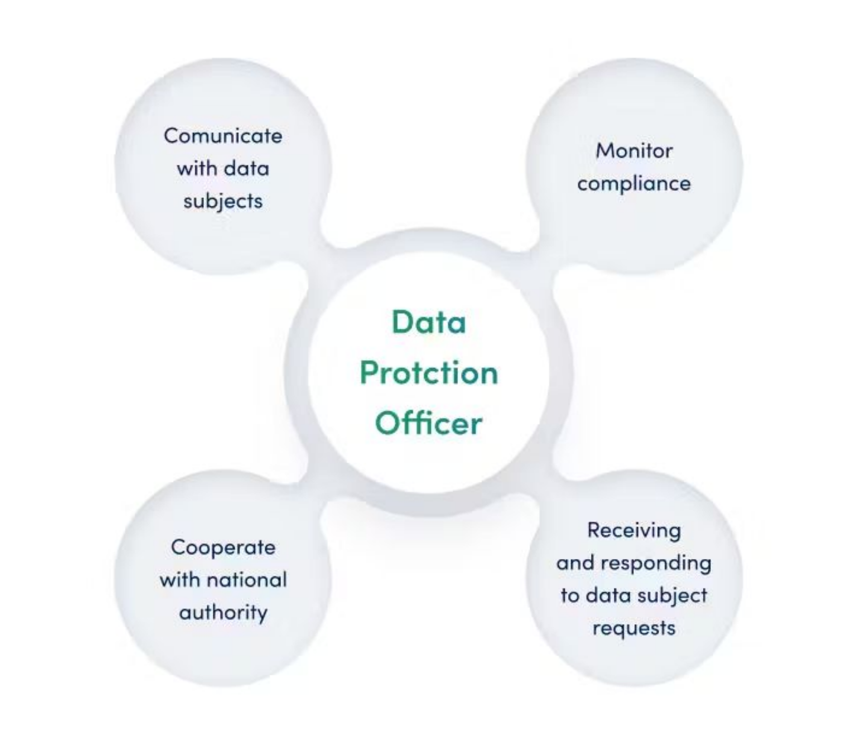
   a. the pseudonymisation and encryption of personal data;

   b. the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;

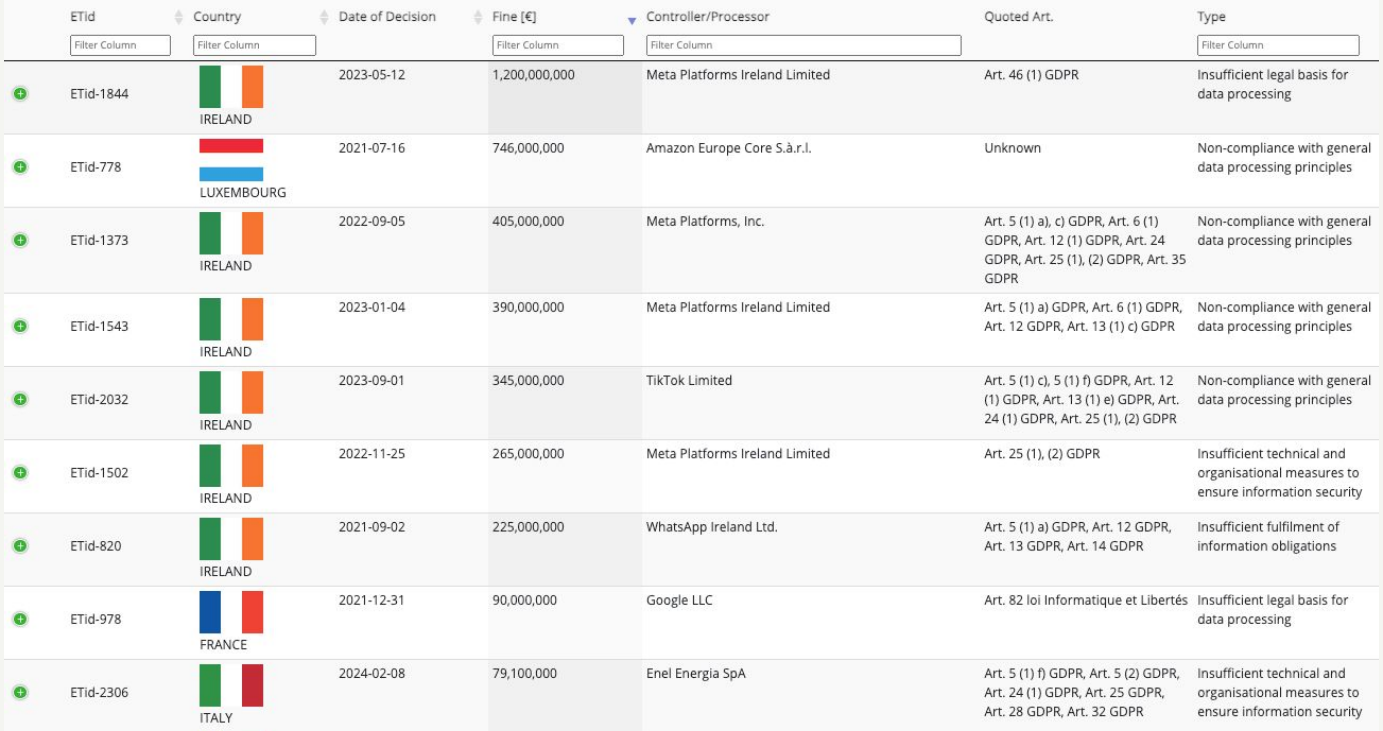
   c. the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;

   d. a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.

Accountability and Supervision

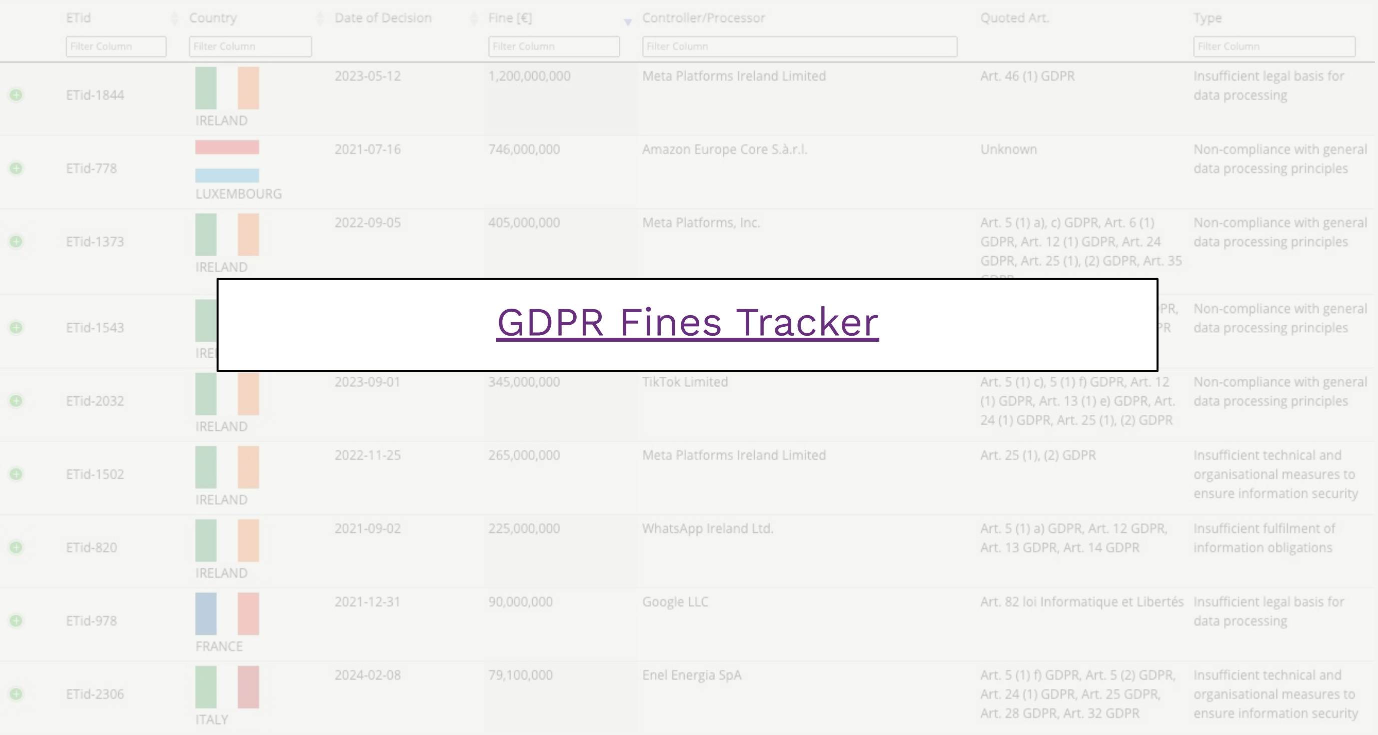

Comunicate with data subjects
Monitor compliance
Data Protction Officer
Cooperate with national authority
Receiving and responding to data subject requests

Accountability and Supervision



	ETid	Country	Date of Decision	Fine [€]	Controller/Processor	Quoted Art.	Type
	Filter Column	Filter Column		Filter Column	Filter Column		Filter Column
⊕	ETid-1844	IRELAND	2023-05-12	1,200,000,000	Meta Platforms Ireland Limited	Art. 46 (1) GDPR	Insufficient legal basis for data processing
⊕	ETid-778	LUXEMBOURG	2021-07-16	746,000,000	Amazon Europe Core S.à.r.l.	Unknown	Non-compliance with general data processing principles
⊕	ETid-1373	IRELAND	2022-09-05	405,000,000	Meta Platforms, Inc.	Art. 5 (1) a), c) GDPR, Art. 6 (1) GDPR, Art. 12 (1) GDPR, Art. 24 GDPR, Art. 25 (1), (2) GDPR, Art. 35 GDPR	Non-compliance with general data processing principles
⊕	ETid-1543	IRELAND	2023-01-04	390,000,000	Meta Platforms Ireland Limited	Art. 5 (1) a) GDPR, Art. 6 (1) GDPR, Art. 12 GDPR, Art. 13 (1) c) GDPR	Non-compliance with general data processing principles
⊕	ETid-2032	IRELAND	2023-09-01	345,000,000	TikTok Limited	Art. 5 (1) c), 5 (1) f) GDPR, Art. 12 (1) GDPR, Art. 13 (1) e) GDPR, Art. 24 (1) GDPR, Art. 25 (1), (2) GDPR	Non-compliance with general data processing principles
⊕	ETid-1502	IRELAND	2022-11-25	265,000,000	Meta Platforms Ireland Limited	Art. 25 (1), (2) GDPR	Insufficient technical and organisational measures to ensure information security
⊕	ETid-820	IRELAND	2021-09-02	225,000,000	WhatsApp Ireland Ltd.	Art. 5 (1) a) GDPR, Art. 12 GDPR, Art. 13 GDPR, Art. 14 GDPR	Insufficient fulfilment of information obligations
⊕	ETid-978	FRANCE	2021-12-31	90,000,000	Google LLC	Art. 82 loi Informatique et Libertés	Insufficient legal basis for data processing
⊕	ETid-2306	ITALY	2024-02-08	79,100,000	Enel Energia SpA	Art. 5 (1) f) GDPR, Art. 5 (2) GDPR, Art. 24 (1) GDPR, Art. 25 GDPR, Art. 28 GDPR, Art. 32 GDPR	Insufficient technical and organisational measures to ensure information security

Accountability and Supervision



	ETid	Country	Date of Decision	Fine [€]	Controller/Processor	Quoted Art.	Type
	Filter Column	Filter Column		Filter Column	Filter Column		Filter Column
⊕	ETid-1844	IRELAND	2023-05-12	1,200,000,000	Meta Platforms Ireland Limited	Art. 46 (1) GDPR	Insufficient legal basis for data processing
⊕	ETid-778	LUXEMBOURG	2021-07-16	746,000,000	Amazon Europe Core S.à.r.l.	Unknown	Non-compliance with general data processing principles
⊕	ETid-1373	IRELAND	2022-09-05	405,000,000	Meta Platforms, Inc.	Art. 5 (1) a), c) GDPR, Art. 6 (1) GDPR, Art. 12 (1) GDPR, Art. 24 GDPR, Art. 25 (1), (2) GDPR, Art. 35	Non-compliance with general data processing principles
⊕	ETid-1543	IRE				PR, PR	Non-compliance with general data processing principles
⊕	ETid-2032	IRELAND	2023-09-01	345,000,000	TikTok Limited	Art. 5 (1) c), 5 (1) f) GDPR, Art. 12 (1) GDPR, Art. 13 (1) e) GDPR, Art. 24 (1) GDPR, Art. 25 (1), (2) GDPR	Non-compliance with general data processing principles
⊕	ETid-1502	IRELAND	2022-11-25	265,000,000	Meta Platforms Ireland Limited	Art. 25 (1), (2) GDPR	Insufficient technical and organisational measures to ensure information security
⊕	ETid-820	IRELAND	2021-09-02	225,000,000	WhatsApp Ireland Ltd.	Art. 5 (1) a) GDPR, Art. 12 GDPR, Art. 13 GDPR, Art. 14 GDPR	Insufficient fulfilment of information obligations
⊕	ETid-978	FRANCE	2021-12-31	90,000,000	Google LLC	Art. 82 loi Informatique et Libertés	Insufficient legal basis for data processing
⊕	ETid-2306	ITALY	2024-02-08	79,100,000	Enel Energia SpA	Art. 5 (1) f) GDPR, Art. 5 (2) GDPR, Art. 24 (1) GDPR, Art. 25 GDPR, Art. 28 GDPR, Art. 32 GDPR	Insufficient technical and organisational measures to ensure information security

GDPR Fines Tracker

All the best for your projects and the pitch day!