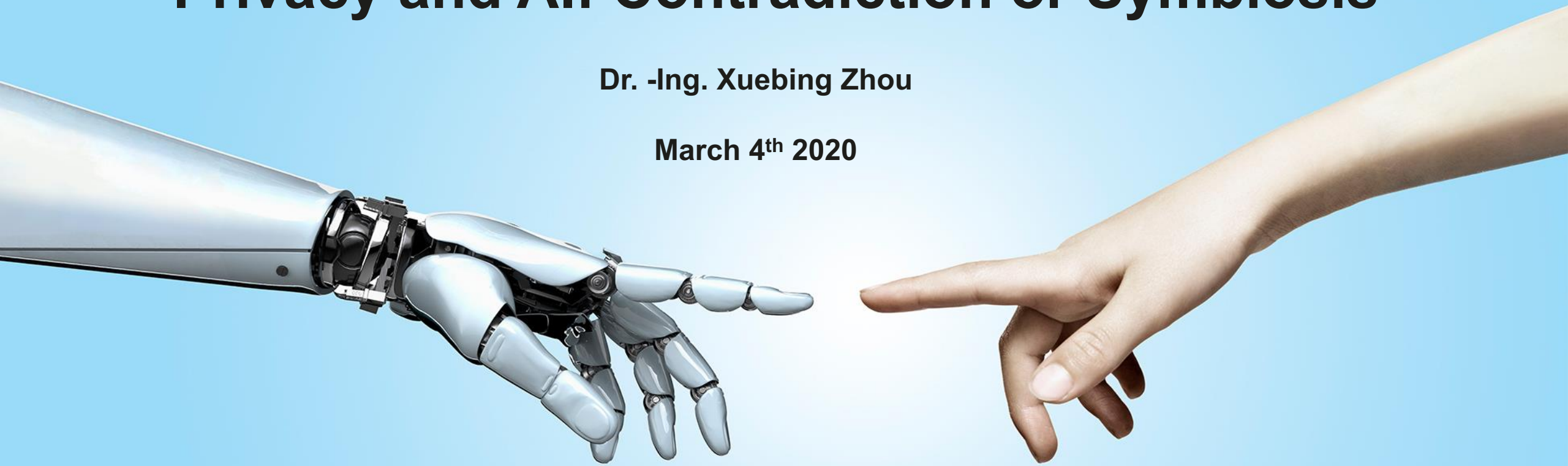
Privacy and AI: Contradiction or Symbiosis

Dr. -Ing. Xuebing Zhou

March 4th 2020



xuebing.zhou@huawei.com
Cyber Security and Privacy Lab, Huawei Technologies

HUAWEI

AI is changing our society considerably

Ethics
Trustworthy,
Fairness
Transparency
Accountability,
Human value
Privacy
Accuracy
Safety
Responsibility
Explainability
Accuracy
Auditability
Adaptability

AI is changing our society considerably

Ethics
Trustworthy,
Fairness
Transparency
Accountability,
Human value

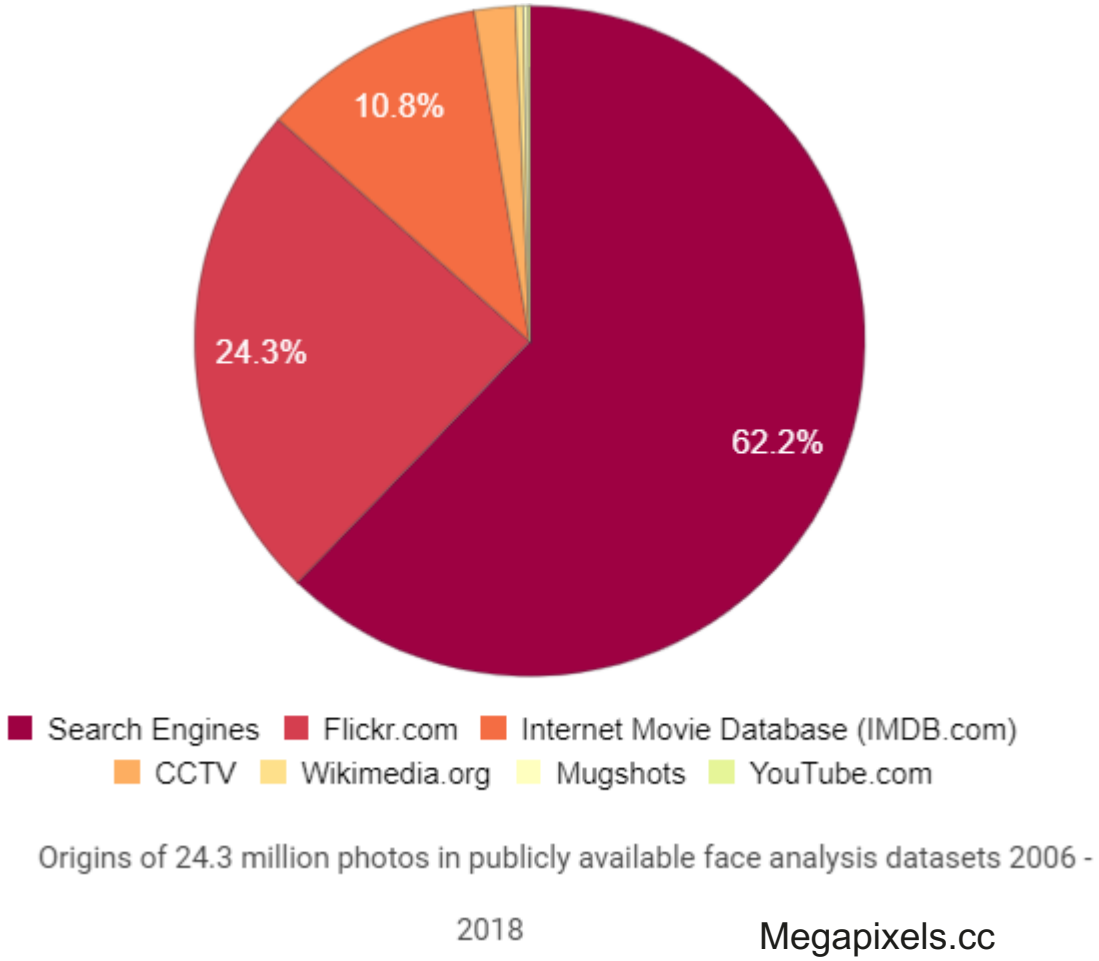
Privacy

Accuracy
Safety
Responsibility
Explainability
Accuracy
Auditability
adaptability

AI and Privacy

Cambridge Analytica

SMART Devices Spy



Origins of 24.3 million photos in publicly available face analysis datasets 2006 - 2018

Megapixels.cc

HUAWEI

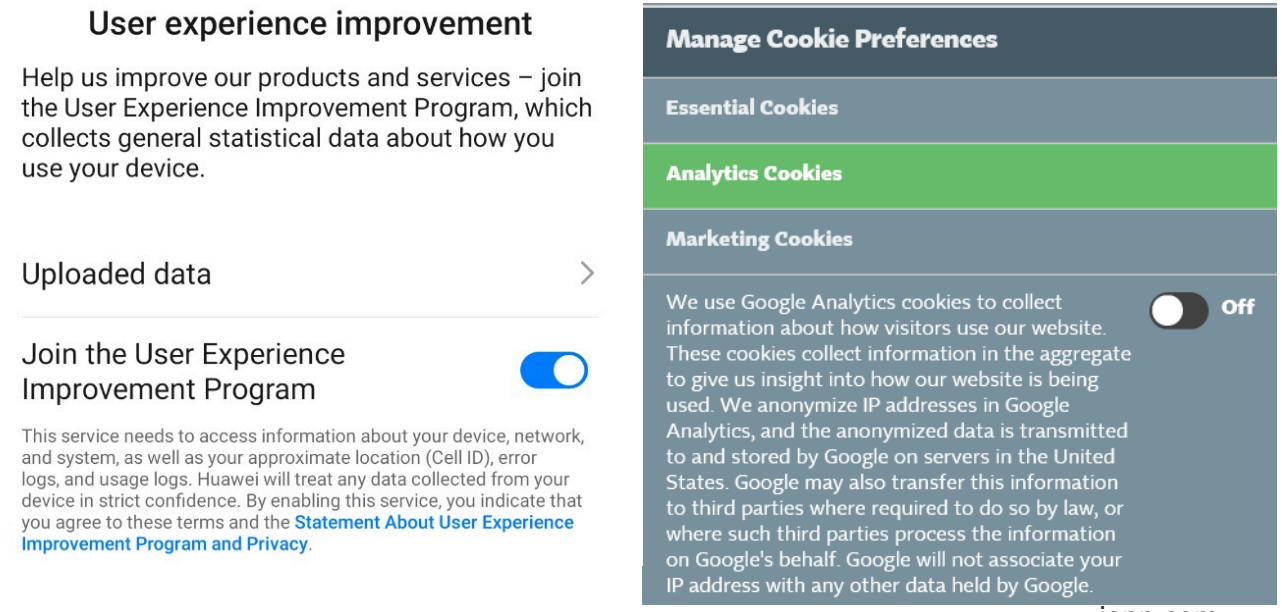
Privacy Protection and AI in Practice

Legal basis assessing personal data

Privacy Policy

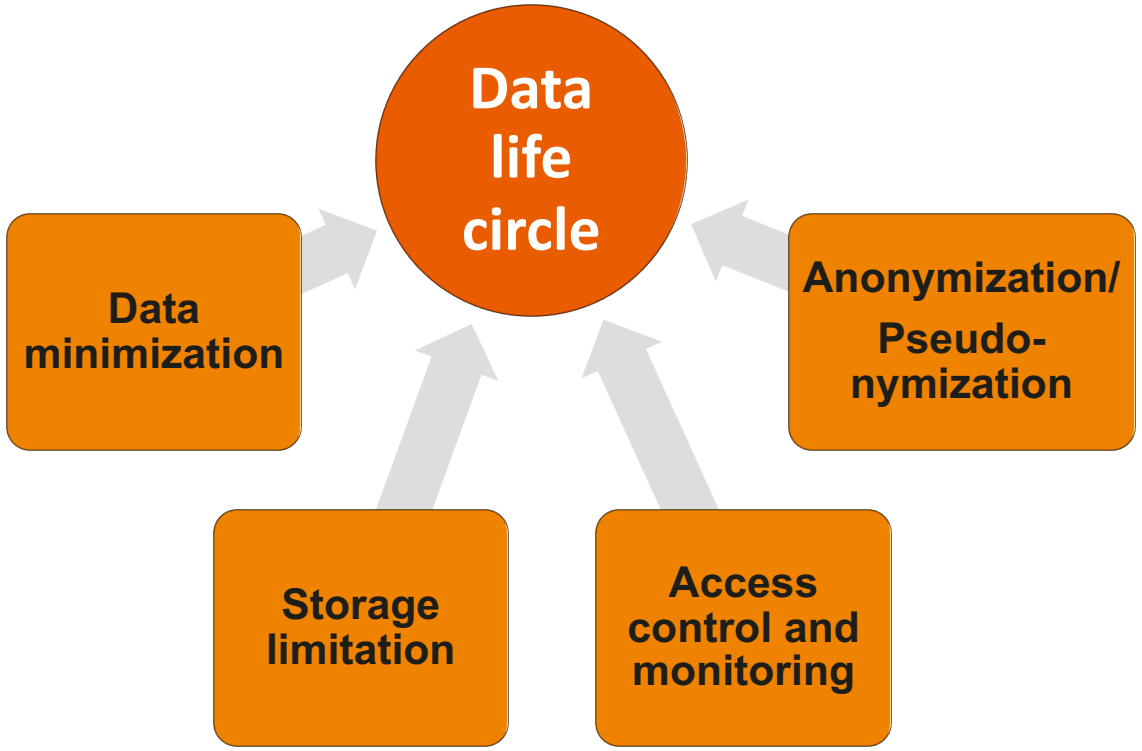
- *Fulfilling your transaction or service requests*
- *Improving our products and services through internal audits, data analysis, and research*

Inform user and leave the choice to the user



iapp.com

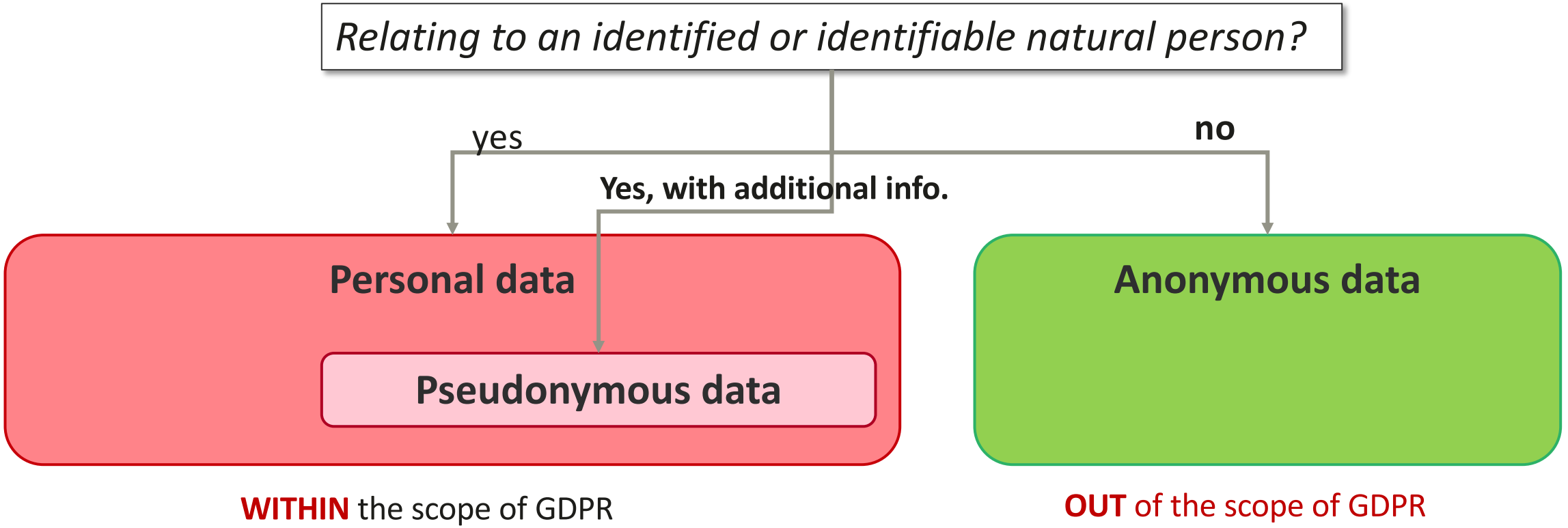
Strong Protection of Personal Data

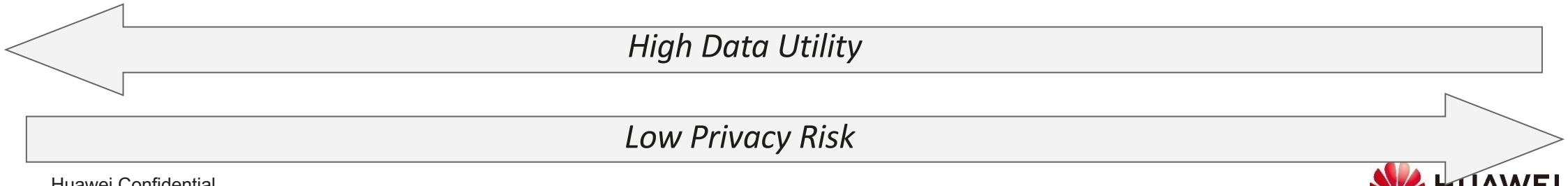


HUAWEI

Anonymization / Pseudonymization

General Data Protection Regulation





Anonymization / Pseudonymization Techniques

Data Masking

- Basic protection on data,
- without consideration on privacy risk

Masking | Truncating | Noise Addition
Enumeration | Permutation | Hashing
Tokenization | Offset

IP:
Masking(10.168.10.1)=10.168.**.*

MAC:
hash(a0a1a2a3a4a5) = 17de9356f8ec

Equivalent class based Anonymization

- De-identify data based on quantitative privacy risk assessment
- Disadvantages: hard to manage dynamic changes of DB, not scalable

k-Anonymity | l-Diversity | δ-Presence | t-Closeness

Sensitive Info.

Name	Phone Nr.	IP Address	Zip	Gender	Age	Place of Birth	Browsing history	# Calls	# SMS	Heart Beat
Peter Scott	168256452	84.72.101.163	60110	Male	30	Frankfurt	spiegel.de, amazon.de, tripadvisor.de, Google.com peter-scott.com	10	2	107
Julie Jason	125678201	54.104.15.181	64512	Female	36	Darmstadt	Google.com, Spiegel.de Amozon.de, msn.de	5	18	80
Paul Richard	182001563	91.96.208.147	65011	Male	38	Augsburg	Google.com, Spiegel.de Amozon.de, ccc.de	0	0	83
Mark Andre	182365472	117.208.184.32	61002	Male	35	Muenchen	Google.com, Spiegel.de Amozon.de, Dailymotion.de	2	5	95

Unique identifiers | Quasi-identifiers | Techn. Data

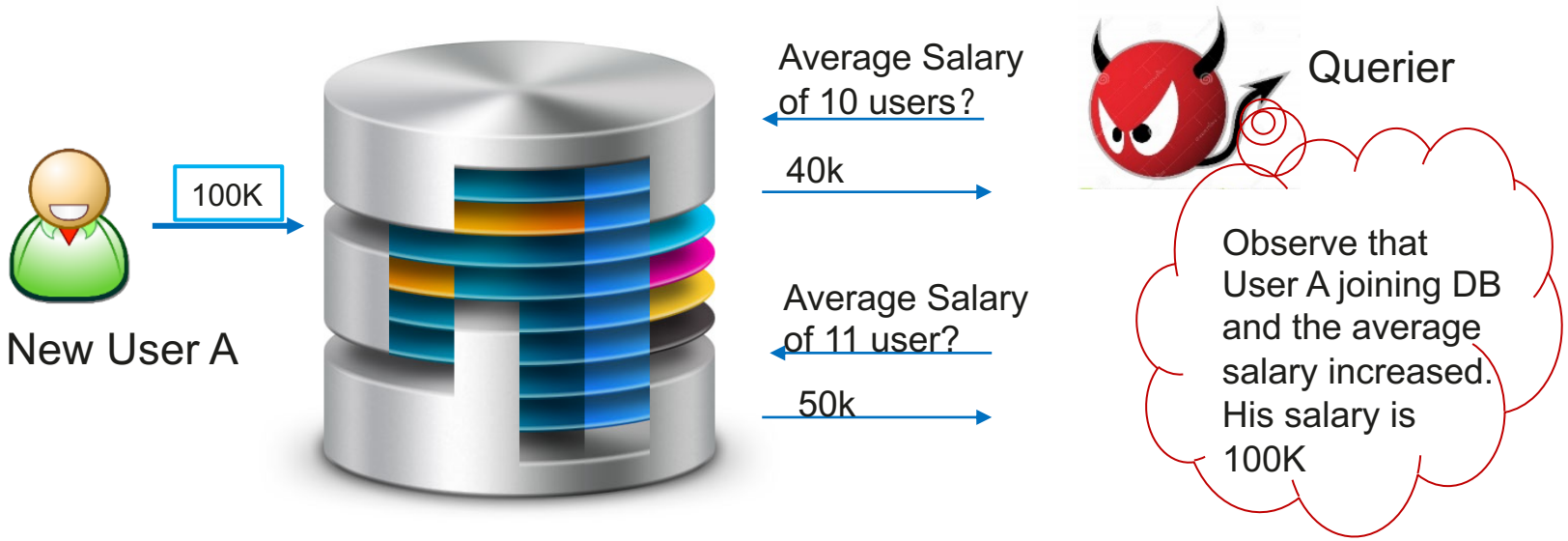
Name	Phone Nr.	IP Address	Zip	Gender	Age	Place of Birth	Browsing history	# Calls	# SMS	
	168******	*	6****	*	30-40	Hessen	Google.com, Spiegel.de, Amozon.de, msn.de	5-10	5-10	
	168******	*	6****	*	30-40	Hessen	Google.com, Spiegel.de, Amozon.de, msn.de	0-5	10-20	
	182******	*	6****	*	30-40	Bayern	Google.com, Spiegel.de, Amozon.de,	0-5	0-5	
	182******	*	6****	*	30-40	Bayern	Google.com, Spiegel.de, Amozon.de,	0-5	0-5	

k=2, the risk of re-identifying a user is 50%

Anonymization / Pseudonymization Techniques

Differential Privacy

Background:



- Existing information can be combined with other background information and new sensitive information can be derived.

* Differential Privacy, C. Dwork, 2008, https://personal.utdallas.edu/~muratk/courses/privacy08f_files/differential-privacy.pdf

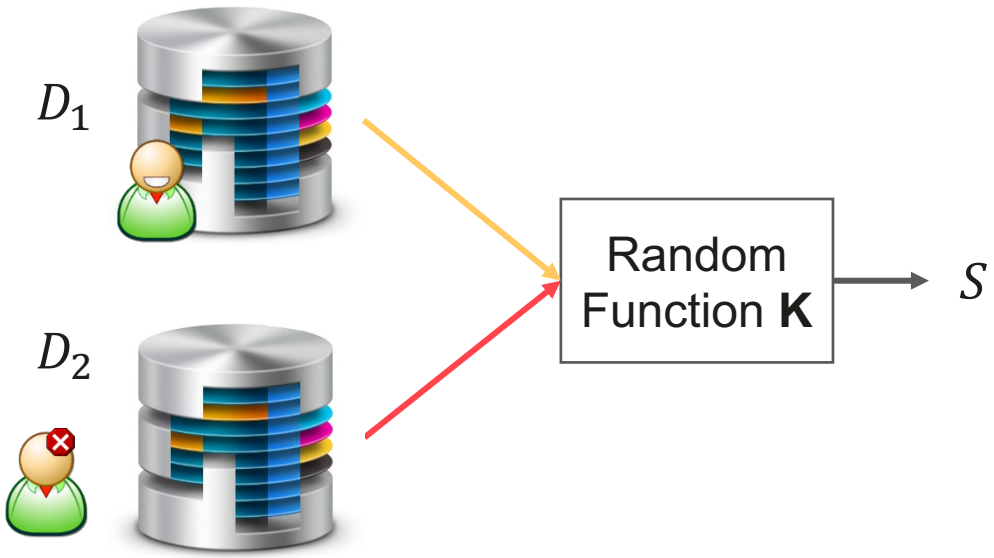
Anonymization / Pseudonymization Techniques

Differential Privacy

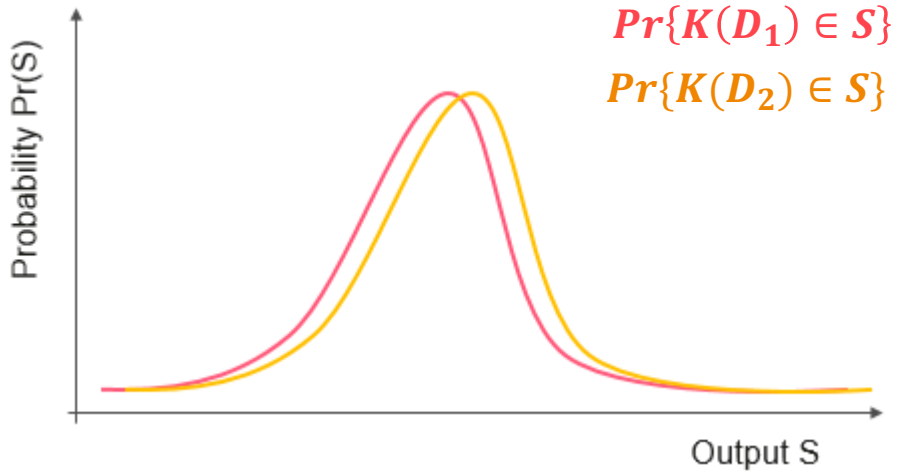
In 2006 Cynthia Dwork, et al. proposed differential privacy, a privacy model to measure privacy risk:

A randomized function *K* gives ε-differential privacy if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq Range(K)$,

$$\Pr\{K(D_1) \in S\} \leq (x + a)^n = e^{\varepsilon} \times \Pr\{K(D_2) \in S\}$$



D_1 and D_2 differs only one Entry

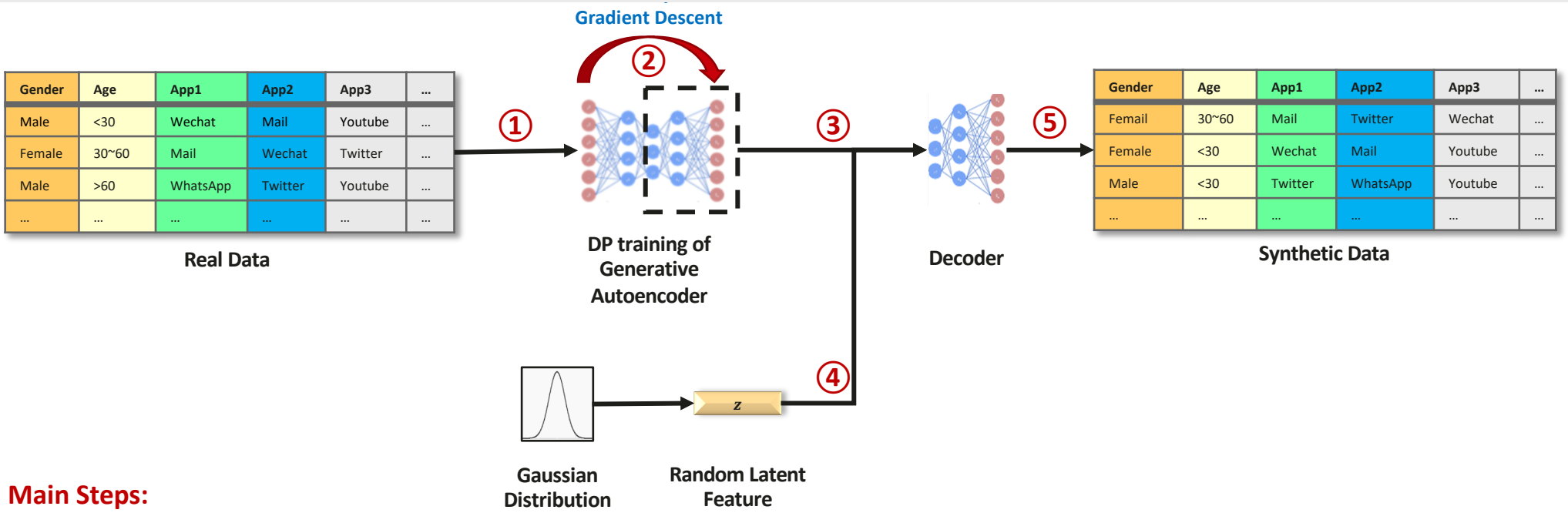


Ratio of distributation of the output bounded by e^{ε}

HUAWEI

Anonymization / Pseudonymization Techniques

Differential Privacy



Main Steps:

① Train the ***generative autoencoder*** with real data

② Apply ***differential privacy*** during model training

③ Extract the ***decoder*** from the trained autoencoder

④ Sample ***random latent feature*** from Gaussian distribution

⑤ Feed latent features into the decoder to ***generate synthetic data***

$\epsilon = 5$

Dataset	Categorical Dimensionality*	Domain Size	Number of Classes	$Accuracy_{real}$	$Accuracy_{generate}$	Accuracy Reduction
Binary10	11	2^{58}	2	95.96%	95.93%	0.03%
Binary20	21	2^{114}	2	93.85%	93.48%	0.37%
Mushroom	23	2^{52}	2	96.51%	94.47%	2.04%
Chess	37	2^{39}	2	95.13%	91.89%	3.24%
Soybean	36	2^{55}	19	90.87%	85.76%	5.11%
Audio	70	2^{82}	24	82.30%	75.62%	6.68%

Other Privacy Attacks on AI

Model Inversion/Reconstruction Attack

- a person's name or unique identifier, and wishes to produce an image of the person associated with that label (i.e., the victim). [1]

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

Fredrikson et al., CCS 2015

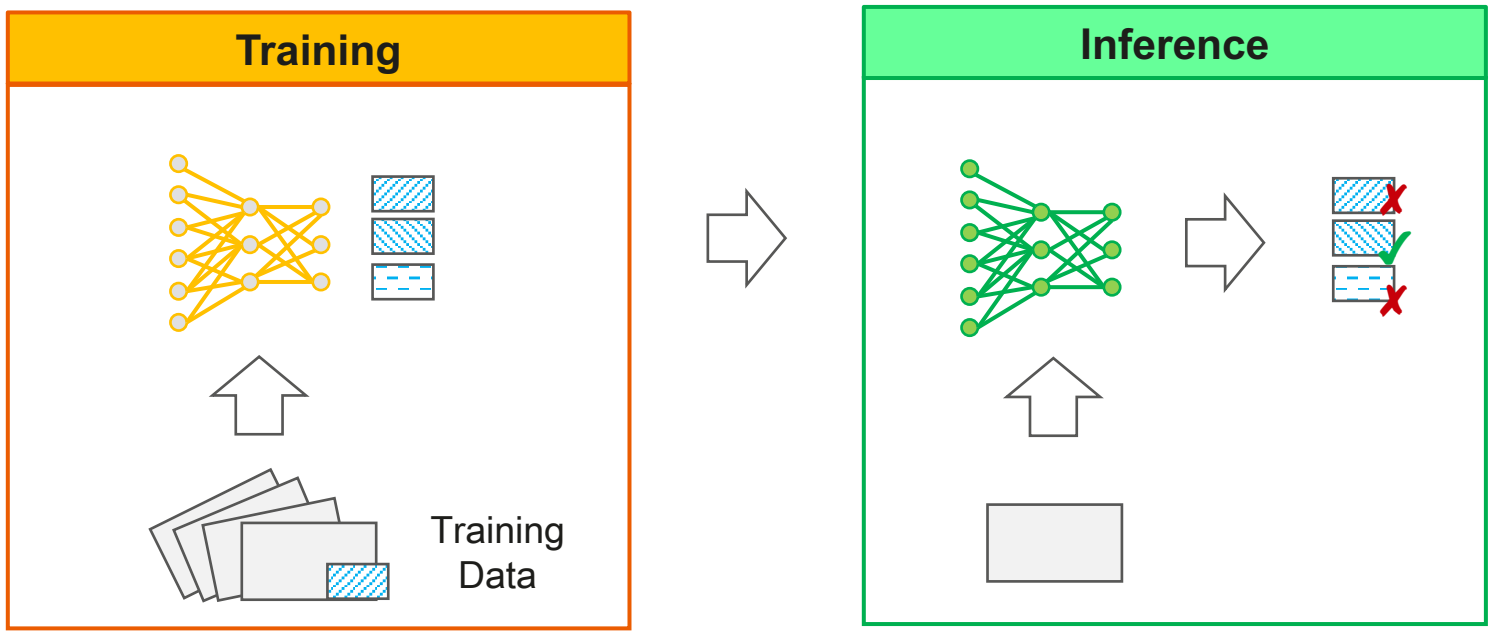
WHEN YOU TRAIN PREDICTIVE MODELS ON INPUT FROM YOUR USERS, IT CAN LEAK INFORMATION IN UNEXPECTED WAYS.[3]

Membership Attack

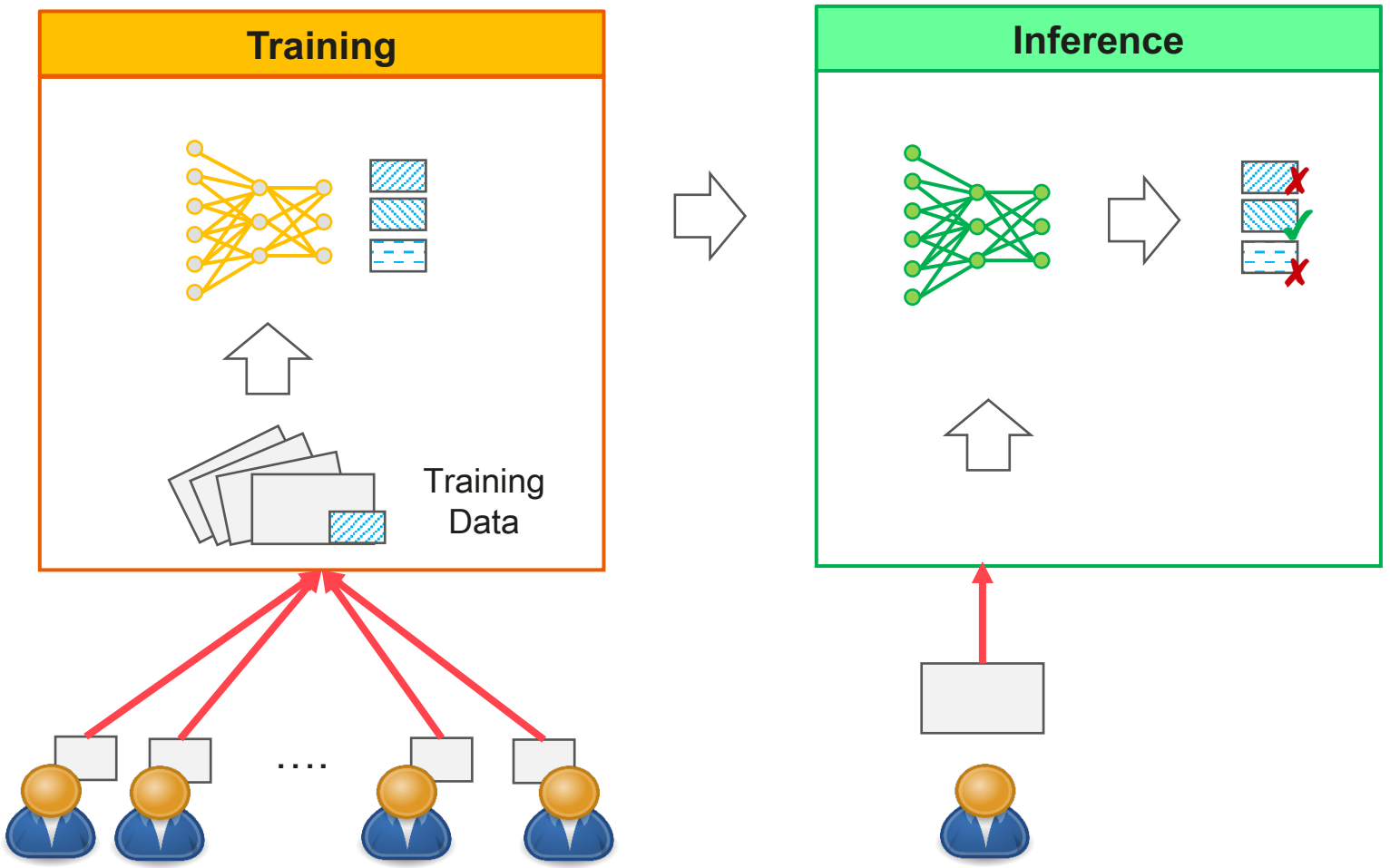
- Shokri. Et. built an attack model and used Google/Amazon ML service as a blackbox. They show that machine learning models leak information about the individual data records on which they were trained. It can be a serious attack if model trained on sensitive data e.g. health data.[2]

[1]Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures.
[2]R. Shokri, M. Stronati, C. Song and V. Shmatikov, "Membership Inference Attacks Against Machine Learning Models," *2017 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, 2017
[3]https://xkcd.com/2169/

Advanced Privacy Protection for AI

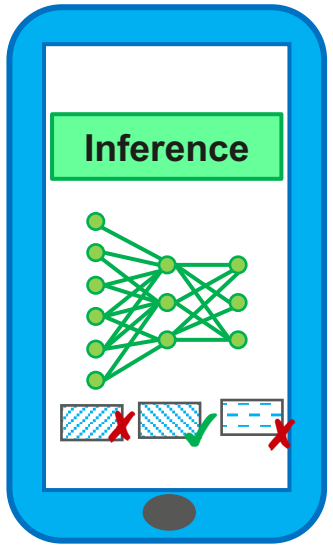


Advanced Privacy Protection for AI



If both training and inference on cloud, user can expose sensitive information to the service provider

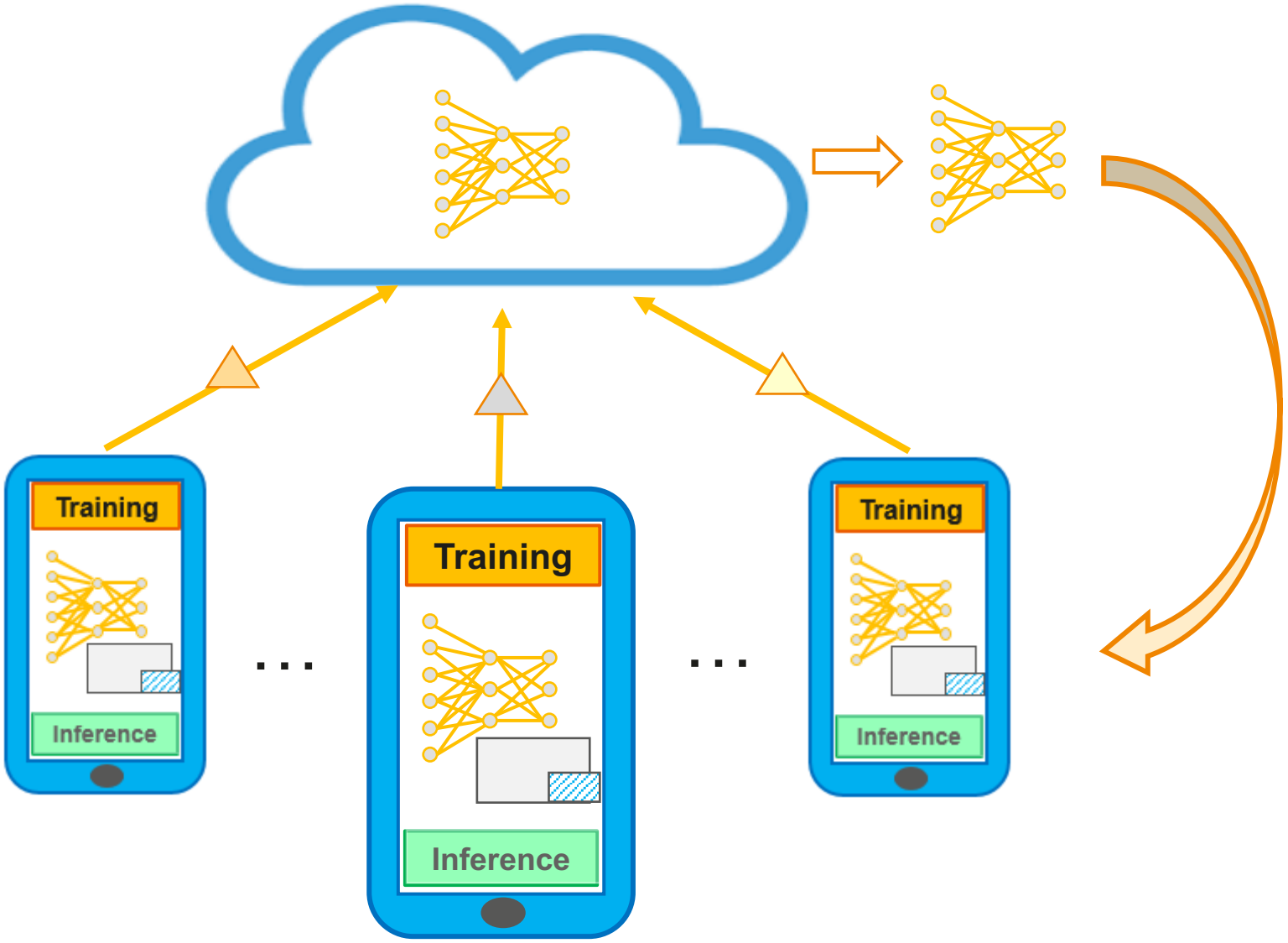
Advanced Privacy Protection for AI



On the device intelligence

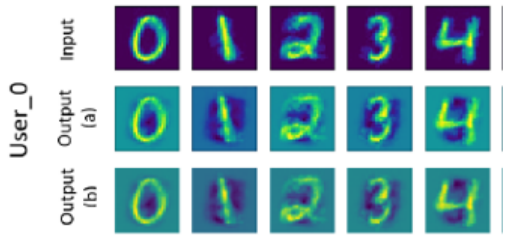
- Inference on the device, no interaction with the service provider, perfect privacy protection
- High requirements on local processing, power consumption, storage of large AI model
- Applied model need be well trained

Advanced Privacy Protection for AI



Federated Learning

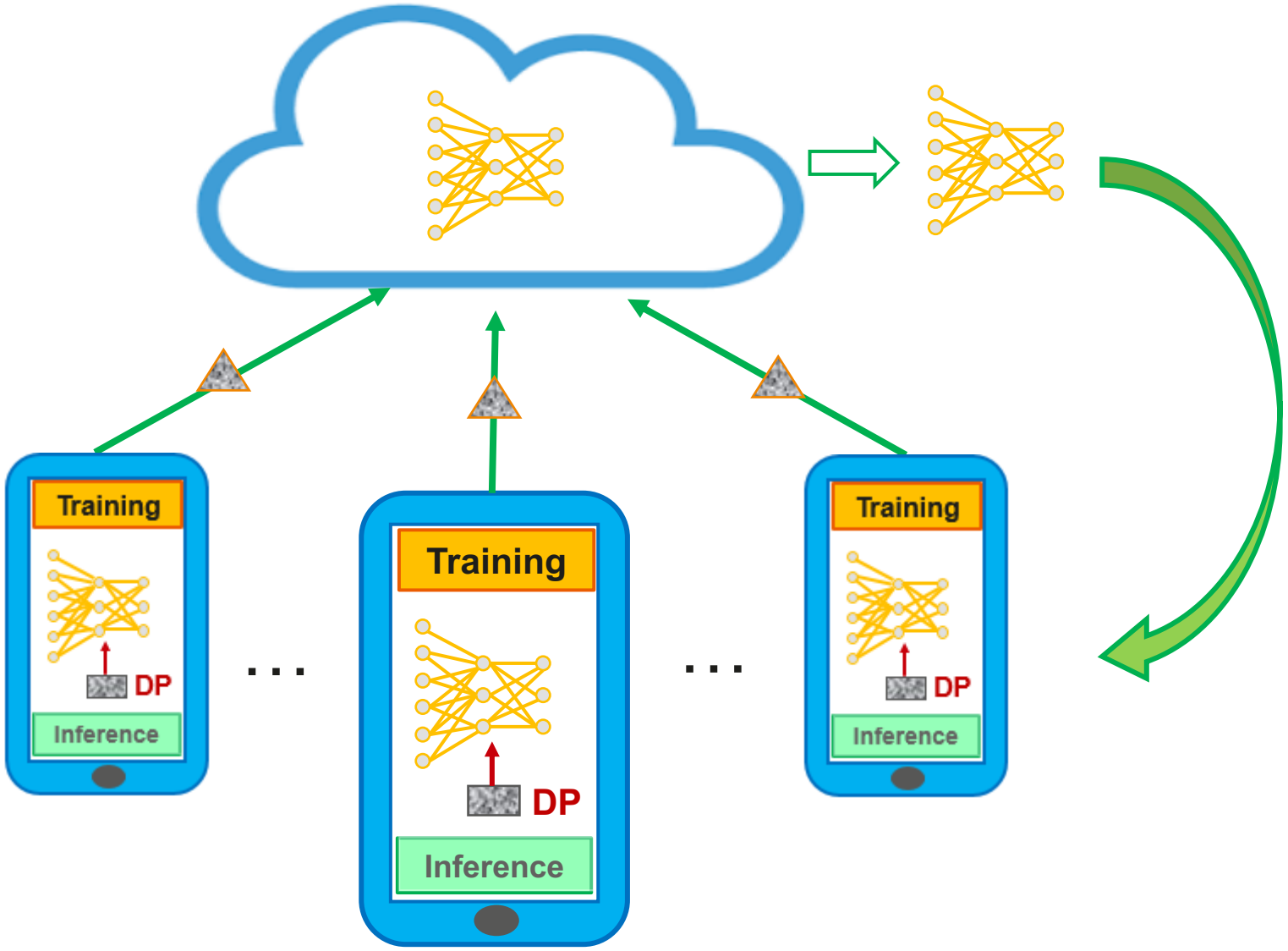
- Model trained locally, only model updates are sent to the server. Server aggregates the model updates and distributed to the clients.
- Privacy is better protected without sharing raw data with server



Reconstruction attack on model update

HUAWEI

Advanced Privacy Protection for AI



Federated Learning+ Differential Privacy

- Add differentially private noise in local training or on the model update
- Prevent that model update reveal information about the local training data
- However, it reduces the accuracy of trained model. Currently it is hard to achieve high privacy for deep learning model

Privacy and AI: Contradiction or Symbiosis

- We cannot gain trust on AI from user without privacy. As an organization, privacy protection need be regulated from management, engineering and technology level.

- Privacy enhancing technologies can solve contradiction between privacy protection and AI
  - Pseudonymization/anonymization techniques provide essential privacy protection in data driven business
  - On-device intelligence, federated learning and differential privacy together can build a comprehensive privacy preserving solution for AI
  - Differential privacy is a very powerful technique and will be standard protection mechanism for AI. However better utility with reasonable privacy protection is needed.

Thank you.

Bring digital to every person, home and organization for a fully connected, intelligent world.