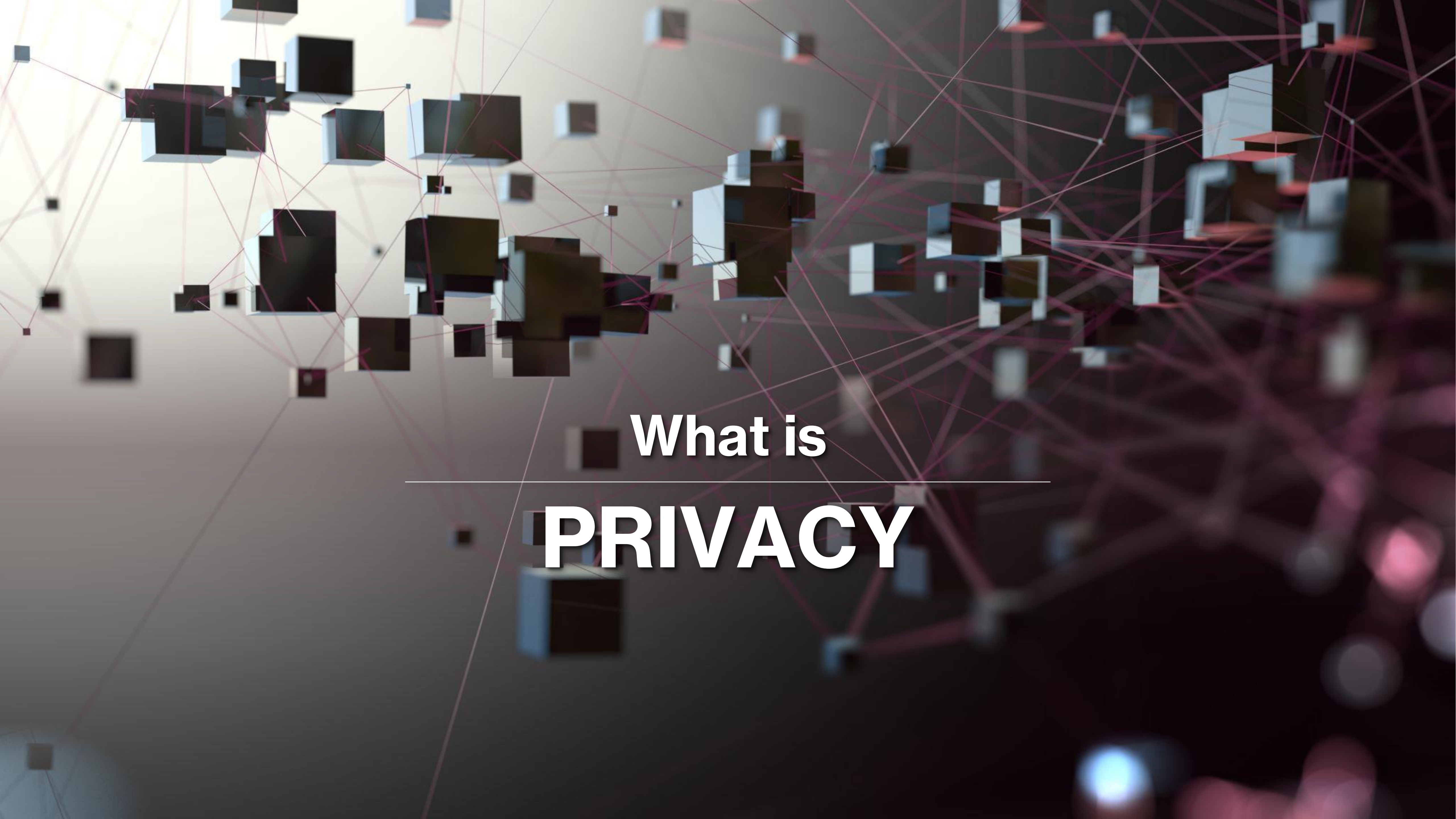
Privacy-Preserving Learning

Eugenio Lomurno
Politecnico di Milano – 18 Feb. 2022
eugenio.lomurno@polimi.it


What is
PRIVACY

“Privacy *is the ability of an* ***individual or group*** *to* ***seclude*** *themselves or* ***information*** *about themselves, and thereby express themselves* ***selectively.”*** - Wikipedia.

General Data Protection Regulation (GDPR)

General Data Protection Regulation (GDPR)

*"The **GDPR** is the **toughest privacy and security law** in the world. It was drafted and passed by the European Union (EU), it **imposes obligations onto organizations** anywhere, so long as they **target or collect data** related to people **in the EU."***

[1] EU General Data Protection Regulation (GDPR):
Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

Europe fit for the Digital Age (21 Apr. 2021)

The Commission proposes today ***new rules*** *and actions aiming to turn Europe into the global hub for* ***trustworthy AI*** *following a risk-based approach*

Europe fit for the Digital Age (21 Apr. 2021)

*The Commission proposes today **new rules** and actions aiming to turn Europe into the global hub for **trustworthy AI** following a risk-based approach*

- ***Unacceptable risk***
- ***High-risk***
- ***Limited risk***
- ***Minimal risk***

Europe fit for the Digital Age (21 Apr. 2021)

*The Commission proposes today **new rules** and actions aiming to turn Europe into the global hub for **trustworthy AI** following a risk-based approach*

- ***Unacceptable risk*** *(e.g. voice assistants encouraging dangerous behaviors, systems allowing social scoring).*
- ***High-risk***
- ***Limited risk***
- ***Minimal risk***

Europe fit for the Digital Age (21 Apr. 2021)

*The Commission proposes today **new rules** and actions aiming to turn Europe into the global hub for **trustworthy AI** following a risk-based approach*

- ***Unacceptable risk*** *(e.g. voice assistants encouraging dangerous behaviors, systems allowing social scoring).*
- ***High-risk*** *(e.g. AI application in robot-assisted surgery, evaluation of the reliability of evidence, scoring of exams).*
- ***Limited risk***
- ***Minimal risk***

Europe fit for the Digital Age (21 Apr. 2021)

*The Commission proposes today **new rules** and actions aiming to turn Europe into the global hub for **trustworthy AI** following a risk-based approach*

*- **Unacceptable risk** (e.g. voice assistants encouraging dangerous behaviors, systems allowing social scoring).*
*- **High-risk** (e.g. AI application in robot-assisted surgery, evaluation of the reliability of evidence, scoring of exams).*
*- **Limited risk** (e.g. chatbots).*
*- **Minimal risk***

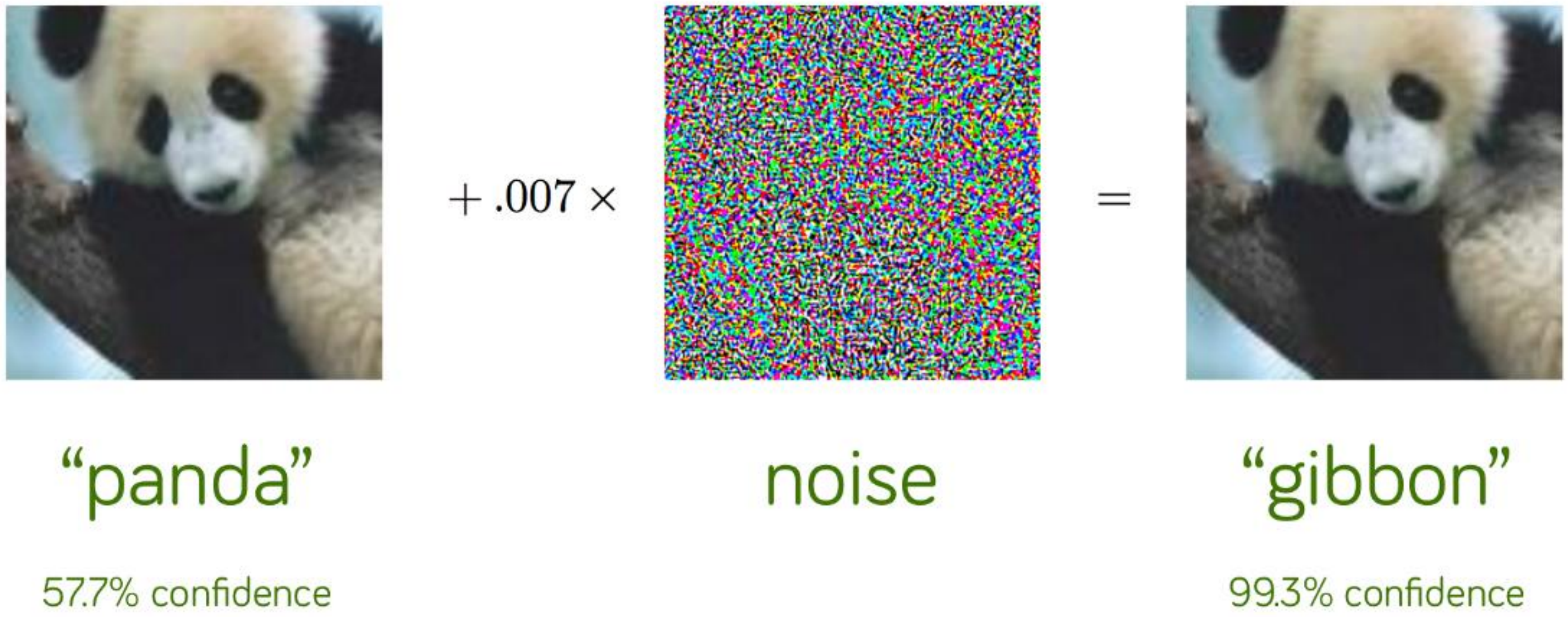
Europe fit for the Digital Age (21 Apr. 2021)

*The Commission proposes today **new rules** and actions aiming to turn Europe into the global hub for **trustworthy AI** following a risk-based approach*

- ***Unacceptable risk*** *(e.g. voice assistants encouraging dangerous behaviors, systems allowing social scoring).*
- ***High-risk*** *(e.g. AI application in robot-assisted surgery, evaluation of the reliability of evidence, scoring of exams).*
- ***Limited risk*** *(e.g. chatbots).*
- ***Minimal risk*** *(e.g. AI-enabled video games, spam filters).*

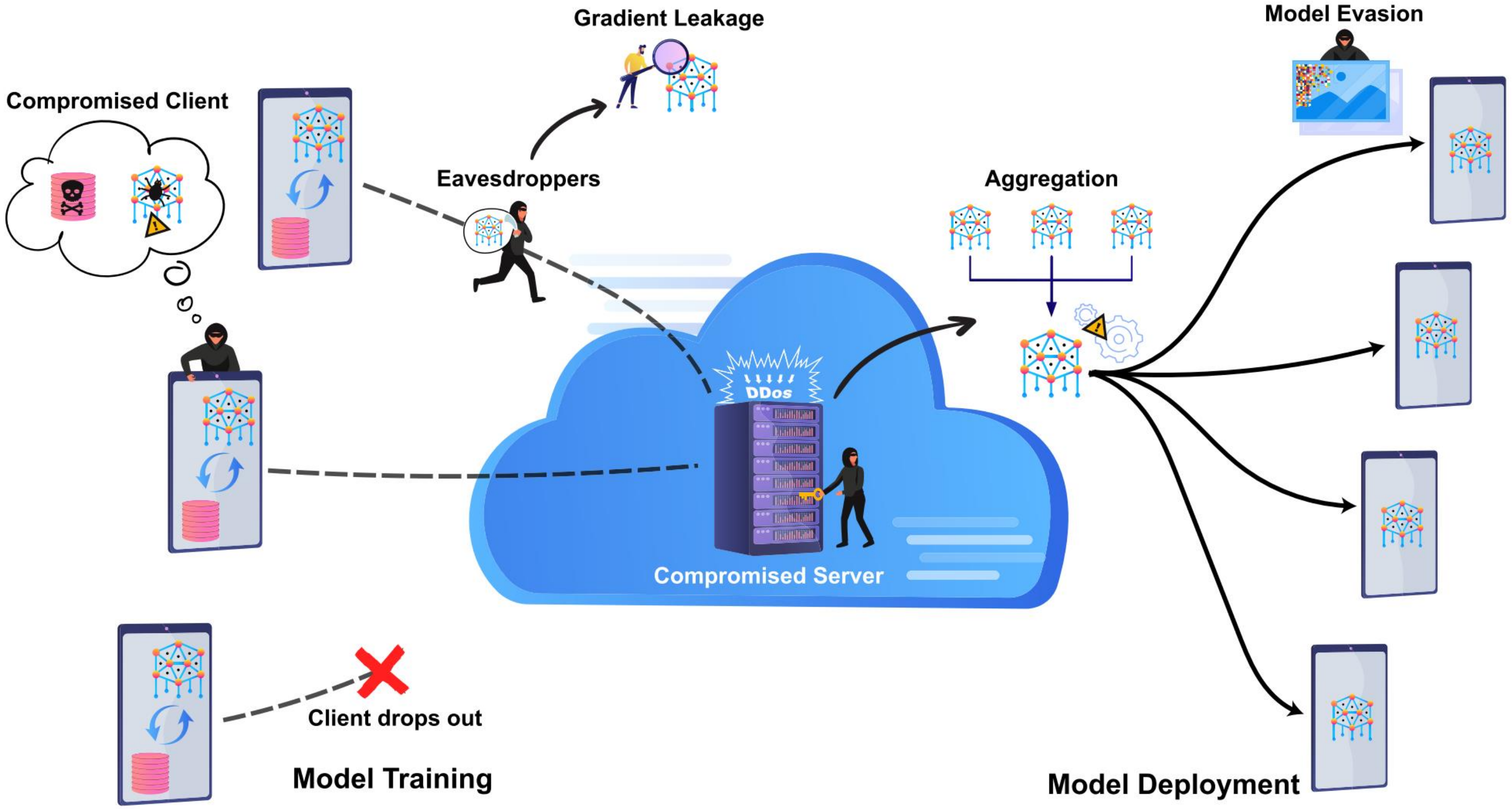
The devil is in the details (and around the corner too)

The devil is in the details (and around the corner too)



[2] Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." arXiv preprint arXiv:1412.6572 (2014).

The devil is in the details (and around the corner too)



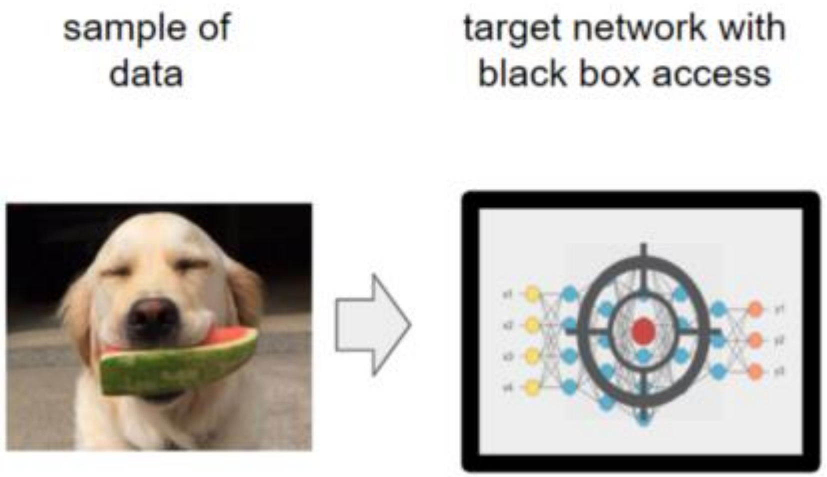
[3] N. Bouacida and P. Mohapatra, “Vulnerabilities in federated learning,” IEEE Access, vol. 9, pp. 63 229–63 249, 2021.

Membership inference attacks

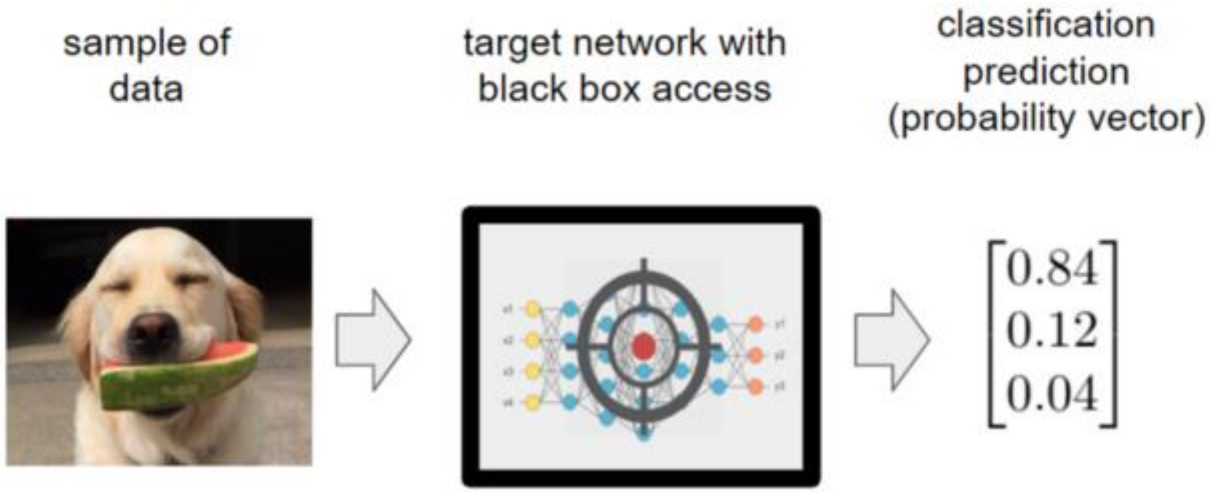
Membership inference attacks

sample of
data

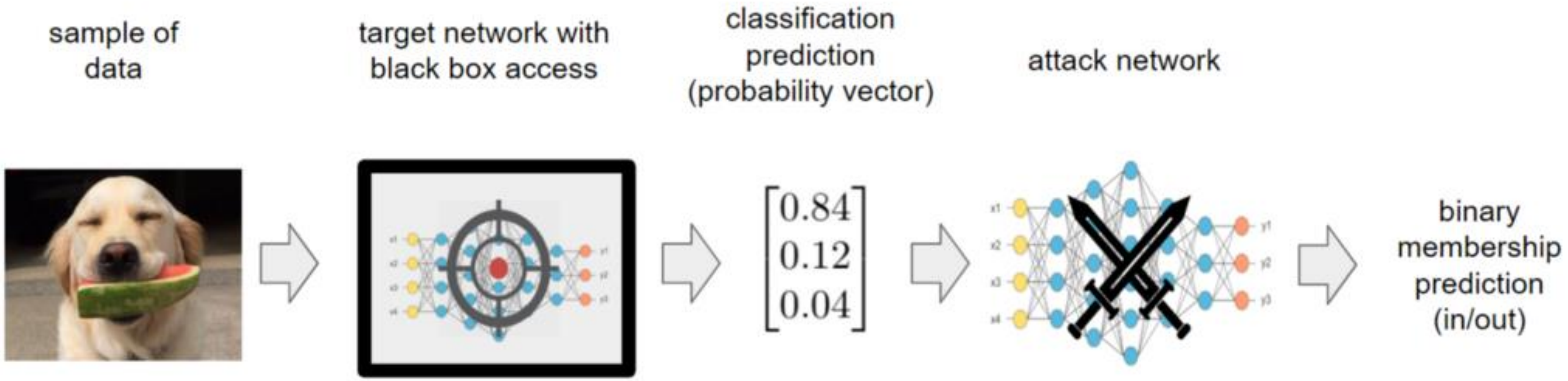
Membership inference attacks



Membership inference attacks

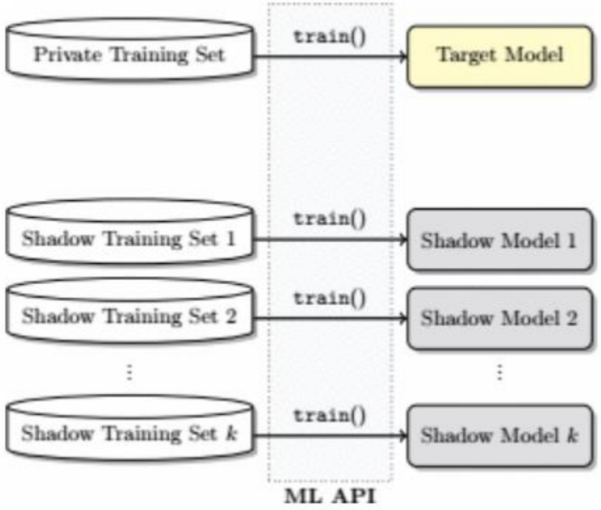


Membership inference attacks

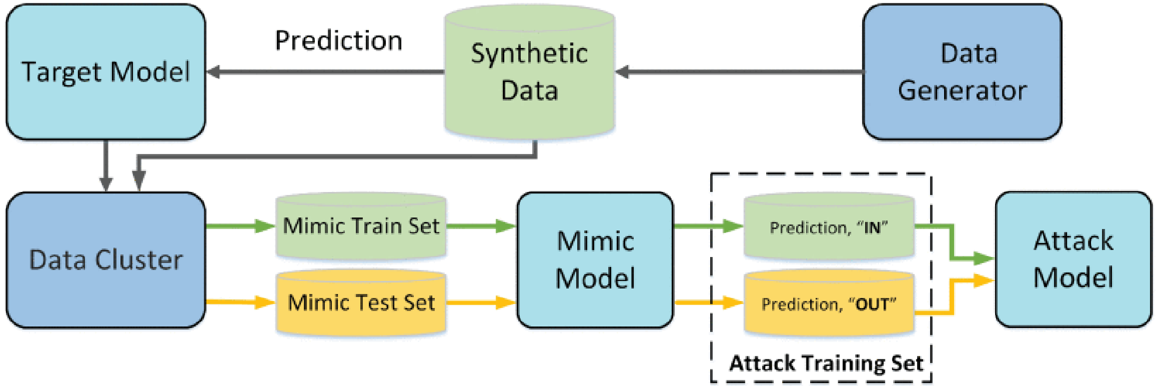


[4] Shokri, Reza, et al. "Membership inference attacks against machine learning models." 2017 IEEE symposium on security and privacy (SP). IEEE, 2017.

Membership inference attacks



[4]



[5]

[4] Shokri, Reza, et al. "Membership inference attacks against machine learning models." 2017 IEEE symposium on security and privacy (SP). IEEE, 2017.
[5] Liu, Gaoyang, et al. "Socinf: Membership inference attacks on social media health data with machine learning."
IEEE Transactions on Computational Social Systems 6.5 (2019): 907-921.

What is

DIFFERENTIAL PRIVACY

*“**Differential privacy** is a system for **publicly sharing information** about a dataset by describing the **patterns of groups** within the dataset while **withholding information** about the **individuals** in the dataset.”* - Wikipedia.

Definition 1
A random mechanism $M: D \rightarrow R$ *with domain* D *and range* R *satisfies* ε*-differential privacy if for any two adjacent inputs* $d, d' \in D$ *and for any subset of outputs* $S \subseteq R$ *it holds that*

$$Pr[M(d) \in S] \leq e^{\varepsilon} Pr[M(d') \in S].$$

[6] Dwork, Cynthia, et al. "Our data, ourselves: Privacy via distributed noise generation." *EUROCRYPT*. Springer, Berlin, Heidelberg, 2006.

Property 1 – (Sequential) *Composability*

Let $M_1, \ldots, M_n$ be n independent random mechanisms whose differential privacy guarantees are $\varepsilon_1, \ldots, \varepsilon_n$, respectively. Then for any function g holds that

$$\varepsilon(g(M_1, \ldots, M_n)) = \sum_{i=1}^{n} \varepsilon_i.$$

If all the components of a mechanism are differentially private, then so is their composition.

[6] Dwork, Cynthia, et al. "Our data, ourselves: Privacy via distributed noise generation." *EUROCRYPT*. Springer, Berlin, Heidelberg, 2006.

Property 2 – Group privacy
A random mechanism $M: D \rightarrow R$ *with domain* D *and range* R *satisfies* ε*-differential privacy if for any two inputs* $d, d' \in D$ *with distance* c *and for any subset of outputs* $S \subseteq R$ *it holds that*

$$Pr[M(d) \in S] \leq e^{\varepsilon c} Pr[M(d') \in S]$$ [7].

[6] Dwork, Cynthia, et al. "Our data, ourselves: Privacy via distributed noise generation." *EUROCRYPT*. Springer, Berlin, Heidelberg, 2006.

Property 3 - Robustness

Given a random mechanism M let F be a deterministic or randomized function defined over the image of M. Then if M satisfies ε-differential privacy, so does $F(M)$.

[6] Dwork, Cynthia, et al. "Our data, ourselves: Privacy via distributed noise generation." *EUROCRYPT*. Springer, Berlin, Heidelberg, 2006.

Definition 2
A random mechanism $M: D \rightarrow R$ *with domain* D *and range* R *satisfies* (ε, δ)*-differential privacy if for any two adjacent inputs* $d, d' \in D$ *and for any subset of outputs* $S \subseteq R$ *it holds that*

$$Pr[M(d) \in S] \leq e^{\varepsilon} Pr[M(d') \in S] + \delta,$$

where $\delta < \frac{1}{|d|}$ *is the possibility that* ε*-differential privacy is broken.*

[6] Dwork, Cynthia, et al. "Our data, ourselves: Privacy via distributed noise generation." *EUROCRYPT*. Springer, Berlin, Heidelberg, 2006.

Gaussian noise mechanism

A common paradigm for approximating a deterministic real-valued function $f: D \to \mathbb{R}$ *with a differentially private mechanism is via additive noise calibrated to* f *sensitivity* $S_f = \max(|f(d) - f(d')|)$ *where* d *and* d' *are two adjacent inputs.*

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Gaussian noise mechanism

The Gaussian noise mechanism is defined as

$$M(d) \triangleq f(d) + \mathcal{N}\left(0, S_f^2\sigma^2\right),$$

where $\mathcal{N}(0, S_f^2\sigma^2)$ is the normal Gaussian distribution with mean 0 and standard deviation $S_f\sigma$.

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Gaussian noise mechanism

The Gaussian noise mechanism is defined as

$$M(d) \triangleq f(d) + \mathcal{N}\left(0, S_f^2\sigma^2\right),$$

where $\mathcal{N}(0, S_f^2\sigma^2)$ is the normal Gaussian distribution with mean 0 and standard deviation $S_f\sigma$.

The analysis of the mechanism can be applied post hoc, and there are infinitely many (ε, δ) pairs that satisfy DP requirements [9].

Due to composition theorems, the mechanism can be iteratively applied in Stochastic Gradient Descent algorithms.

[8] Abadi, Martin, et al. "Deep learning with differential privacy." Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Definition 3

Let $M: D \rightarrow R$ be a randomized mechanism and $d, d' \in D$ a pair of adjacent databases. Let aux denote an auxiliary input. For an outcome $o \in R$, the privacy loss at o is defined as

$$c(o; M, aux, d, d') \triangleq log \frac{\Pr[M(aux, d) = o]}{\Pr[M(aux, d') = o]}.$$

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Definition 4

Let $M: D \to R$ be a randomized mechanism and $d, d' \in D$ a pair of adjacent databases. Let aux denote an auxiliary input. The moments accountant is defined as

$$\alpha_M(\lambda) \triangleq \max_{aux,d,d'} \alpha_M(\lambda; aux, d, d'),$$

where $\alpha_M(\lambda; aux, d, d') \triangleq log\mathbb{E}_{o \backsim M(aux,d)}[e^{\lambda c(o;M,aux,d,d')}]$ is the moment generating function evaluated at value λ.

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Differentially private SGD

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

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Differentially private SGD

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 $\mathbf{g}_t(x_i) \leftarrow \nabla_{\theta_t}\mathcal{L}(\theta_t, x_i)$

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Differentially private SGD

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

$\bar{\mathbf{g}}_t(x_i) \leftarrow \mathbf{g}_t(x_i) / \max\left(1, \frac{\|\mathbf{g}_t(x_i)\|_2}{C}\right)$

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Differentially private SGD

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

$\tilde{\mathbf{g}}_t \leftarrow \frac{1}{L}\sum_i \left(\bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I})\right)$

Descent

$\theta_{t+1} \leftarrow \theta_t - \eta_t \tilde{\mathbf{g}}_t$

[8] Abadi, Martin, et al. "Deep learning with differential privacy."
Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

Differentially private SGD

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

  $\tilde{\mathbf{g}}_t \leftarrow \frac{1}{L}\sum_i(\bar{\mathbf{g}}_t(x_i) + \mathcal{N}(0, \sigma^2 C^2 \mathbf{I}))$

  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.

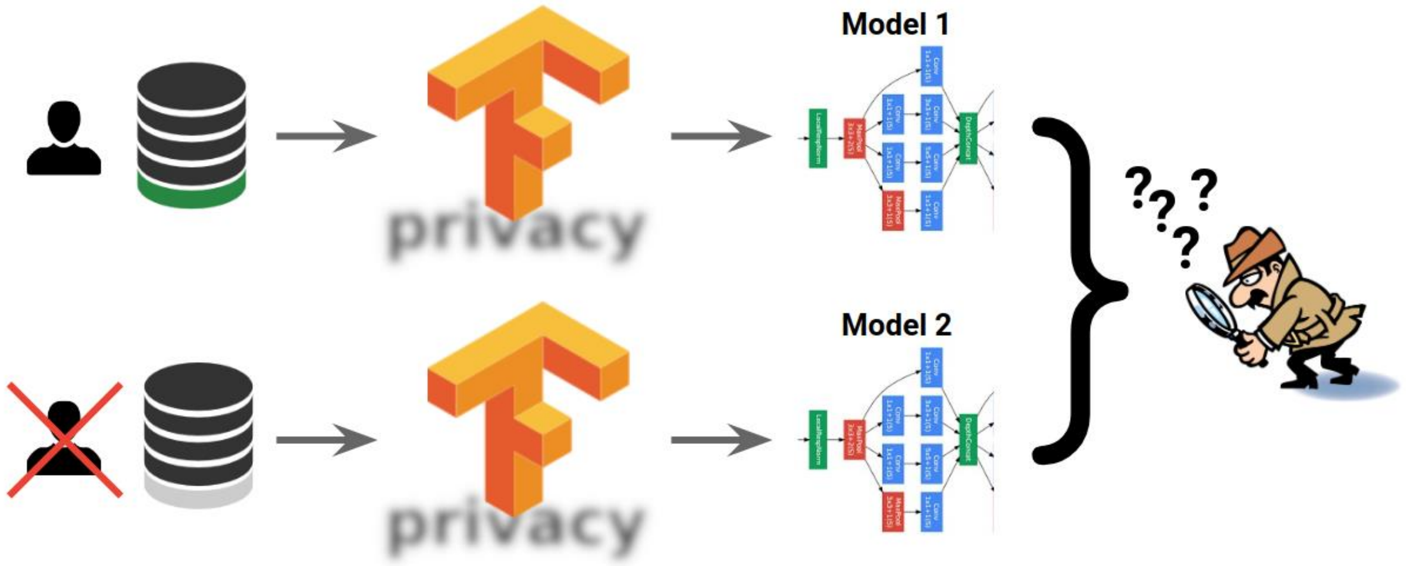
[8] Abadi, Martin, et al. "Deep learning with differential privacy." Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

TensorFlow

TensorFlow Privacy

Train deep learning models using DP Optimizers and vectorized losses.

The privacy analysis is performed in the framework of Rényi Differential Privacy.



[10] Mironov, Ilya. "Rényi differential privacy." 2017 IEEE 30th computer security foundations symposium (CSF). IEEE, 2017.

TensorFlow Privacy

DP-Optimizers take three additional hyperparameters:

Effect of Increasing Hyperparameters On Privacy/Utility/Speed			
Hyperparameter	Privacy	Utility	Speed

TensorFlow Privacy

DP-Optimizers take three additional hyperparameters:

- ***Number of microbatches B** (number of microbatches into which each minibatch is split).*

Effect of Increasing Hyperparameters On Privacy/Utility/Speed			
Hyperparameter	Privacy	Utility	Speed
Number of microbatches B	-	↗	↘

TensorFlow Privacy

DP-Optimizers take three additional hyperparameters:

- ***Number of microbatches*** B *(number of microbatches into which each minibatch is split).*
- ***Clipping norm*** C *(the maximum l2 norm of each individual gradient computed per minibatch).*

Effect of Increasing Hyperparameters On Privacy/Utility/Speed			
Hyperparameter	Privacy	Utility	Speed
Number of microbatches B	-	↗	↘
Clipping norm C	-	?	-

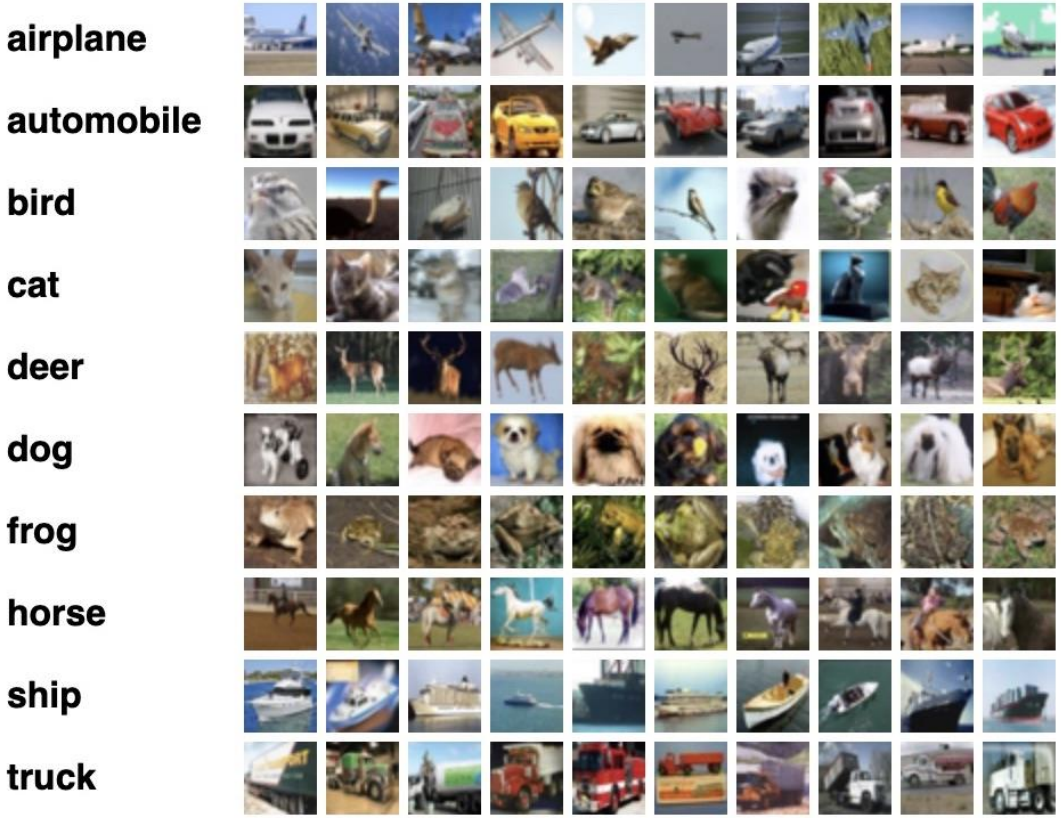
TensorFlow Privacy

DP-Optimizers take three additional hyperparameters:

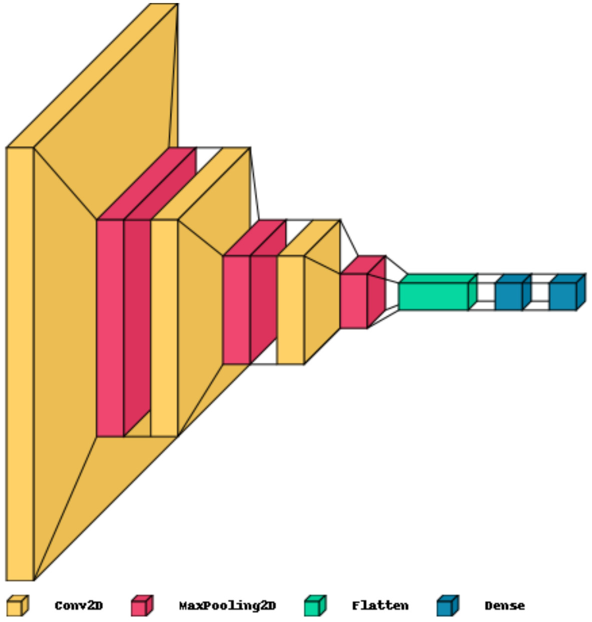
- ***Number of microbatches** B (number of microbatches into which each minibatch is split).*
- ***Clipping norm** C (the maximum l2 norm of each individual gradient computed per minibatch).*
- ***Noise multiplier** σ (ratio of the standard deviation to the clipping norm).*

Effect of Increasing Hyperparameters On Privacy/Utility/Speed			
Hyperparameter	Privacy	Utility	Speed
Number of microbatches B	-	↗	↘
Clipping norm C	-	?	-
Noise multiplier σ	↗	↘	-

A simple example: CIFAR10



CIFAR10 is an open source RGB dataset composed of 60’000 images 32x32 and divided in 10 equally distributed classes.

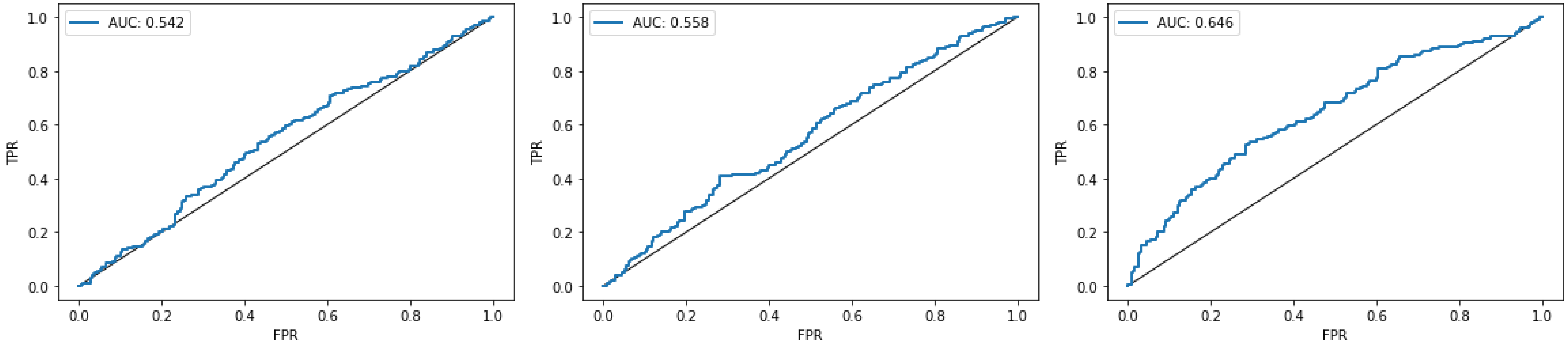


[11] Krizhevsky, Alex, and Geoffrey Hinton. "Learning multiple layers of features from tiny images." (2009).

A simple example: CIFAR10

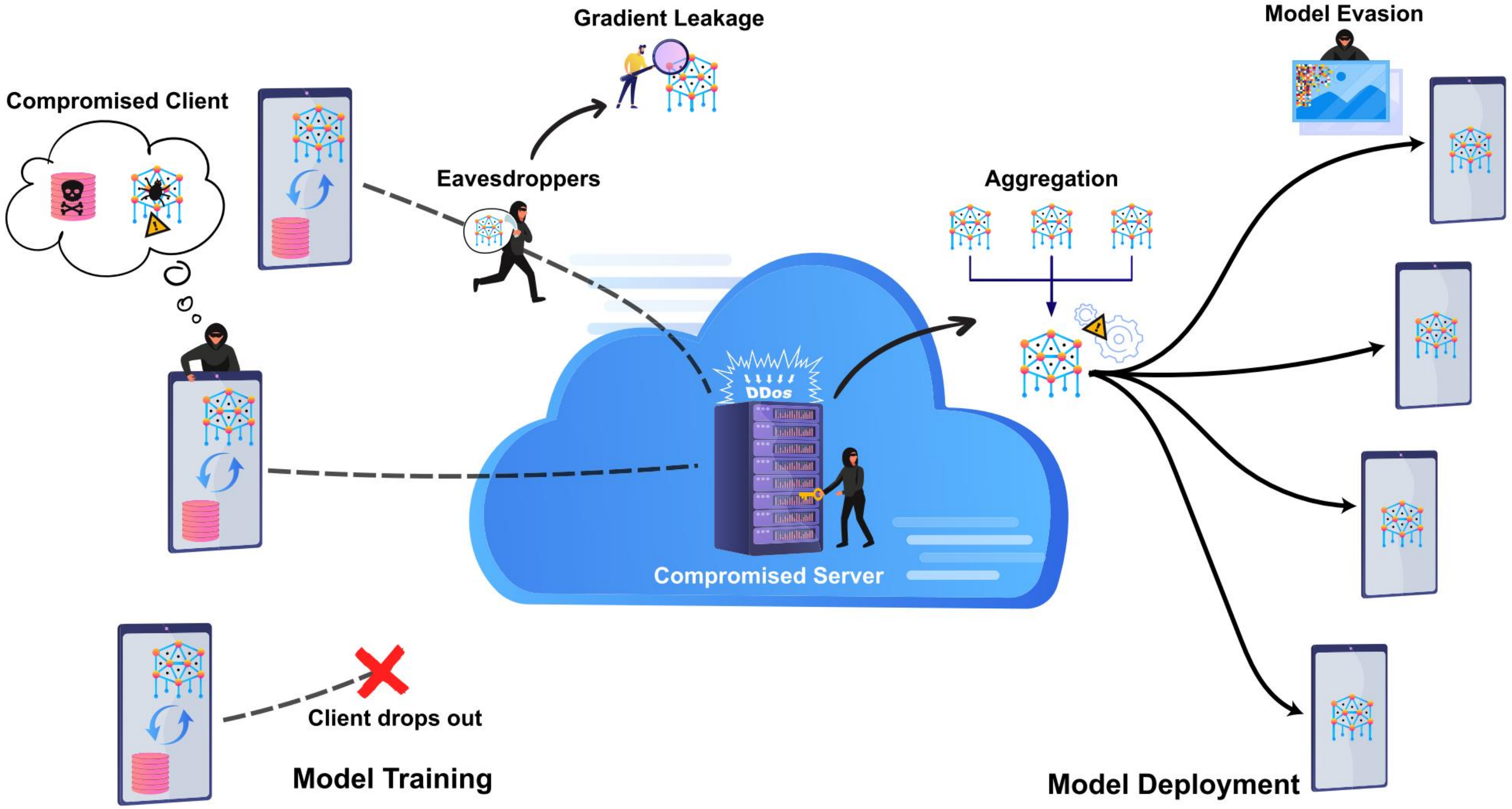
The results of logistic regression-based membership inference attacks

$\varepsilon = 2, \delta = 10^{-5}$ | $\varepsilon = 6, \delta = 10^{-5}$ | *No differential privacy*



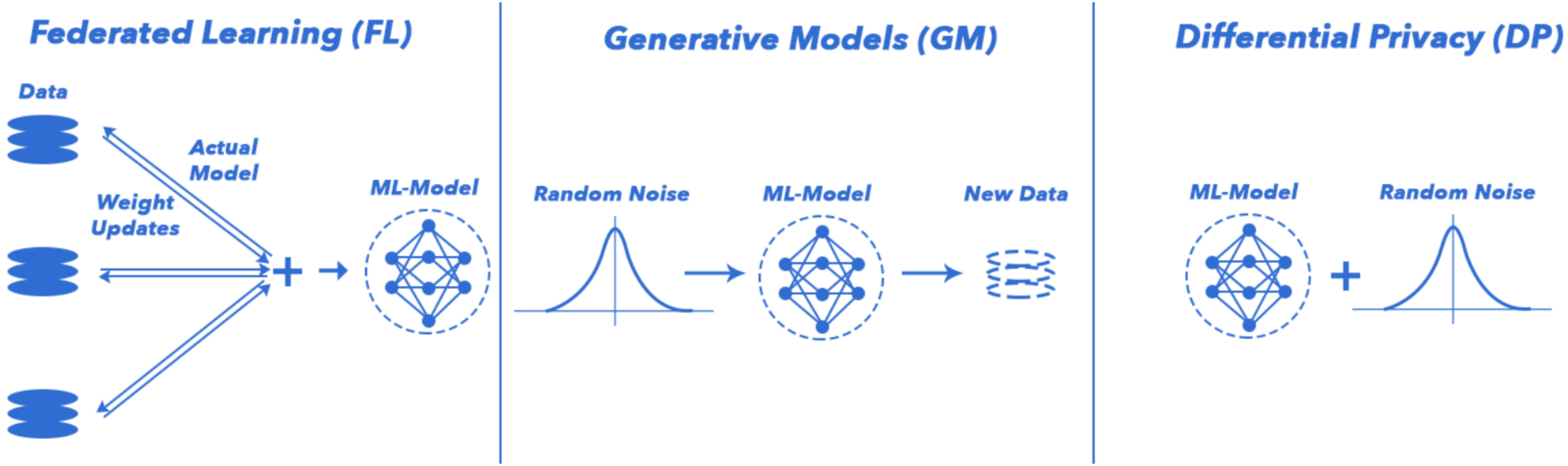
WRAP UP

A world of vulnerabilities...

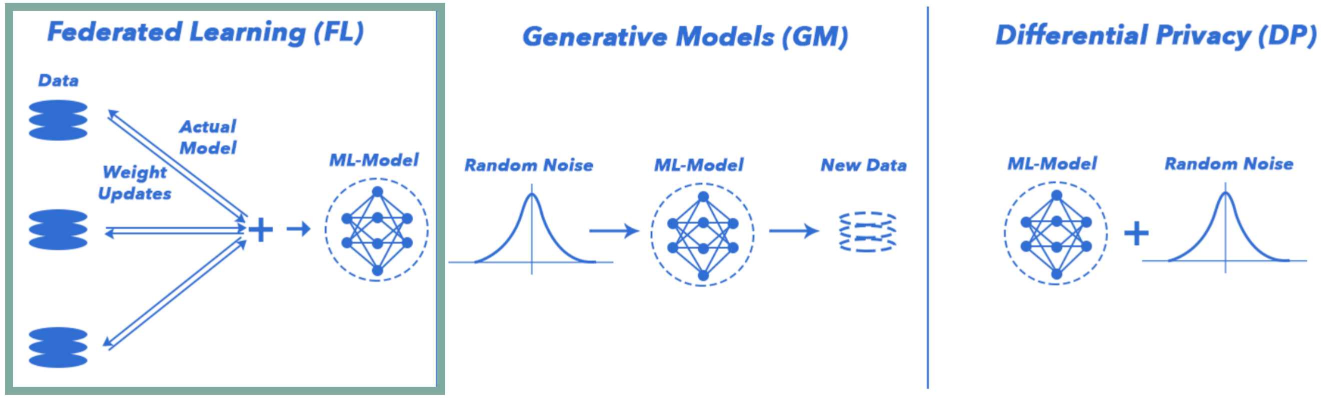


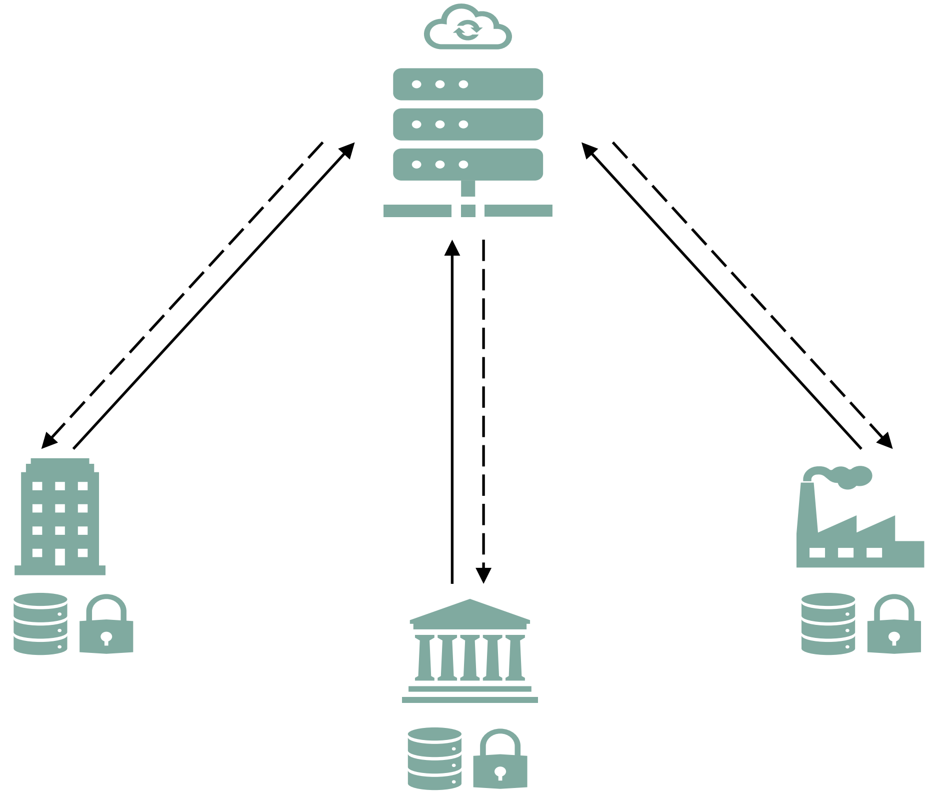
N. Bouacida and P. Mohapatra, “Vulnerabilities in federated learning,” IEEE Access, vol. 9, pp. 63 229–63 249, 2021.

The ingredients



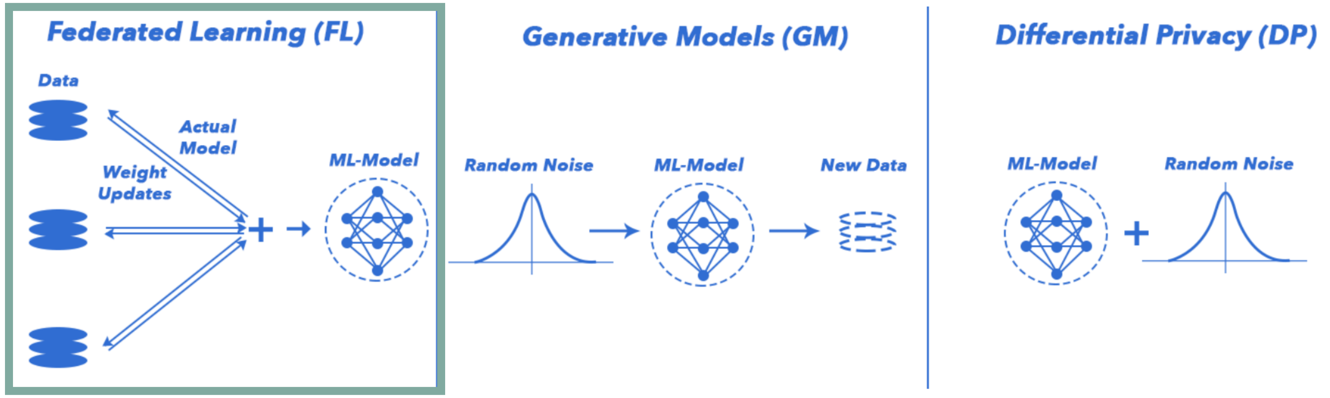
The ingredients

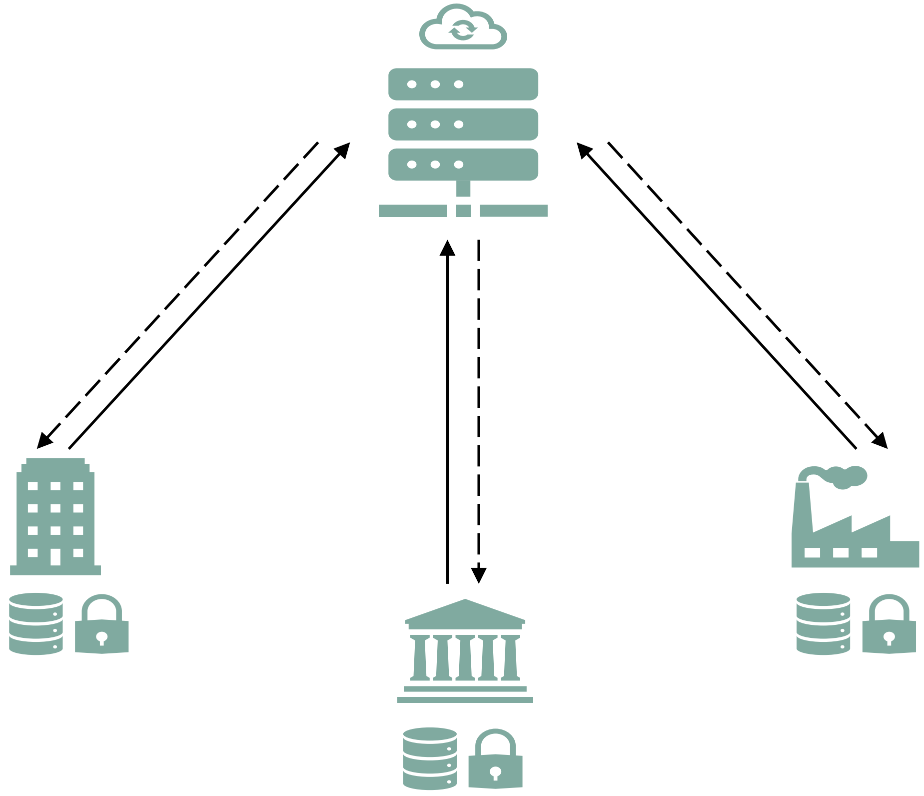




- *Cross-Silo Federated Learning Network composed of 20 members.*

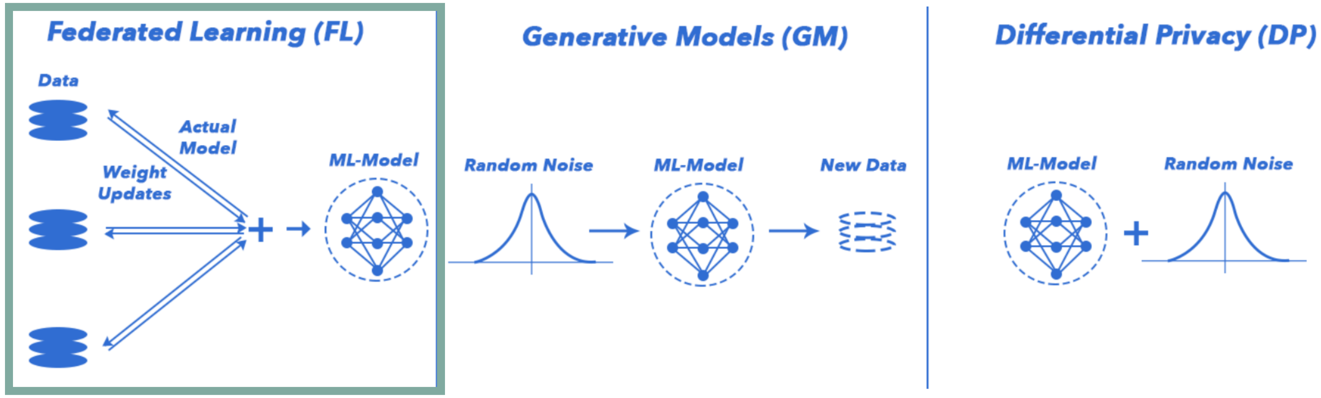
The ingredients

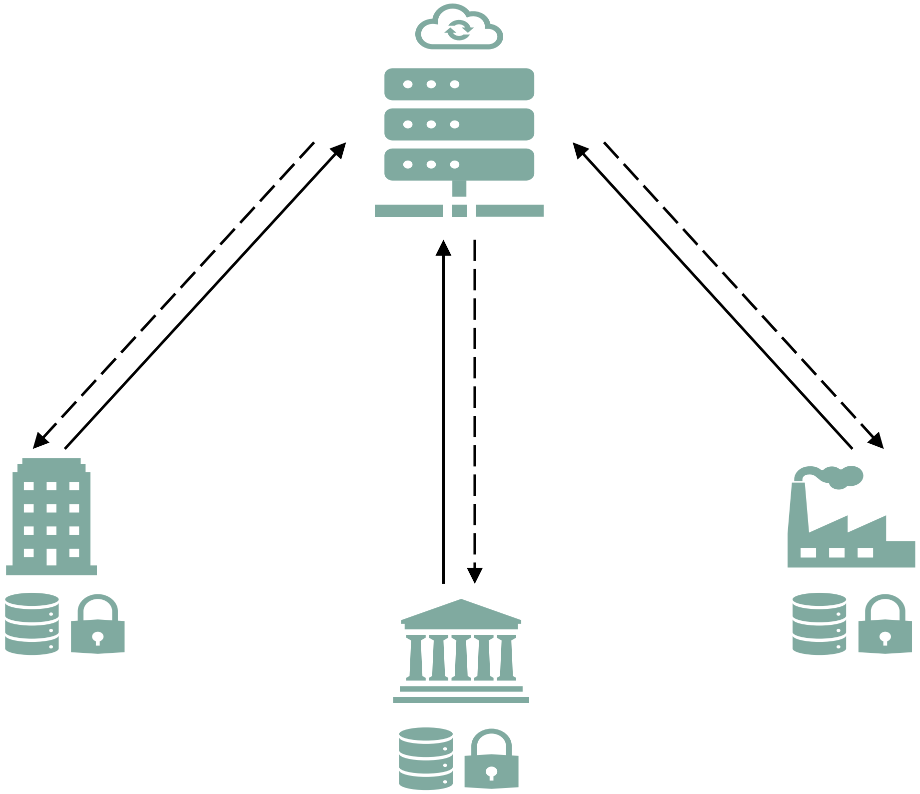




- *Cross-Silo Federated Learning Network composed of 20 members.*
- *7 private equally distributed datasets (tabular/images, balanced/unbalanced).*

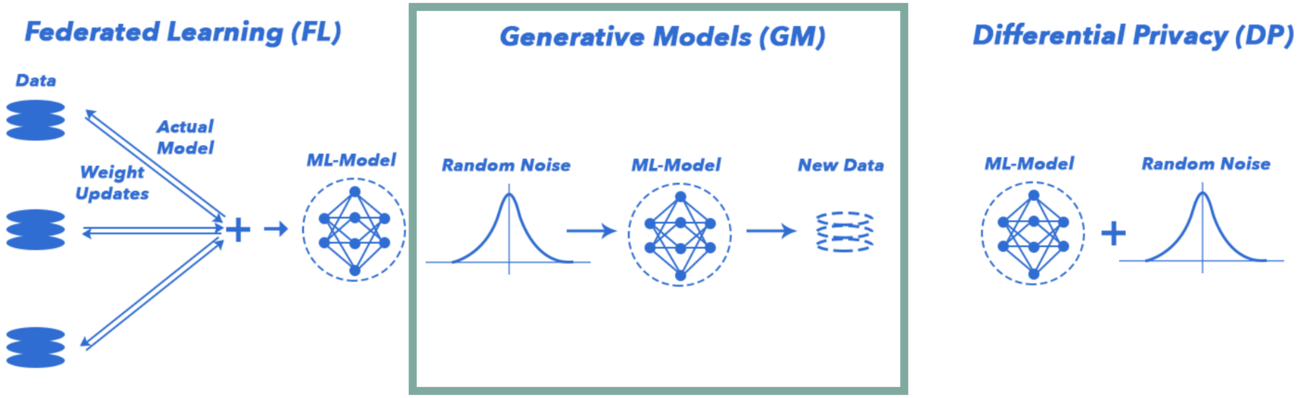
The ingredients

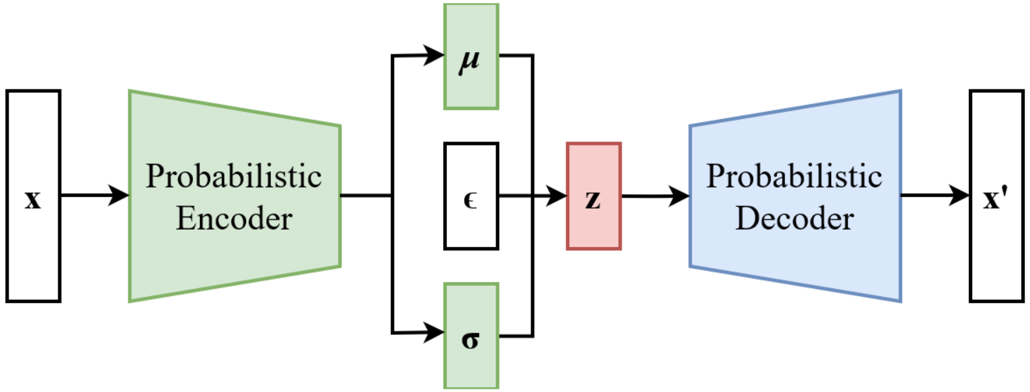




- *Cross-Silo Federated Learning Network composed of 20 members.*
- *7 private equally distributed datasets (tabular/images, balanced/unbalanced).*
- *Hypothesis: all the network's members want to share their own knowledge preserving individuals' privacy.*

The ingredients

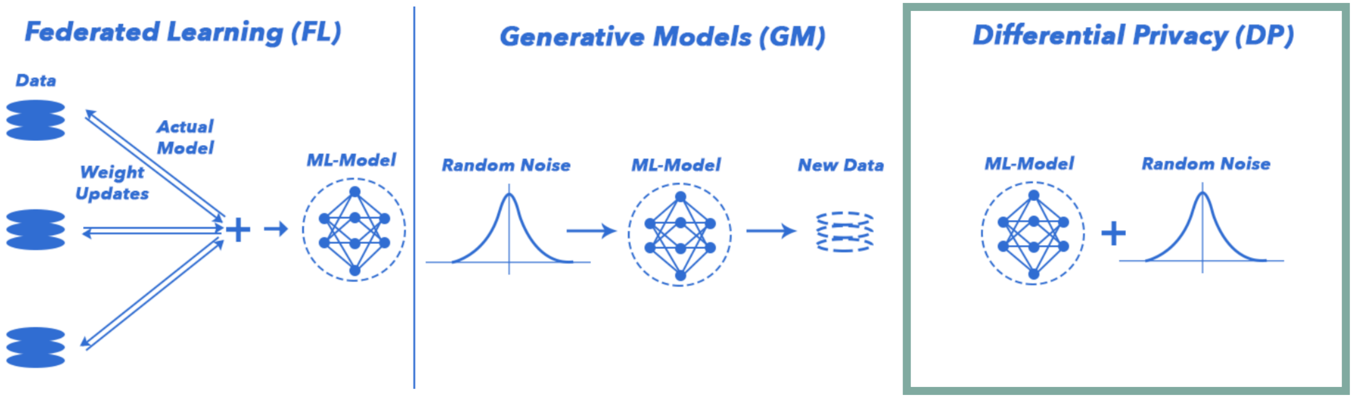


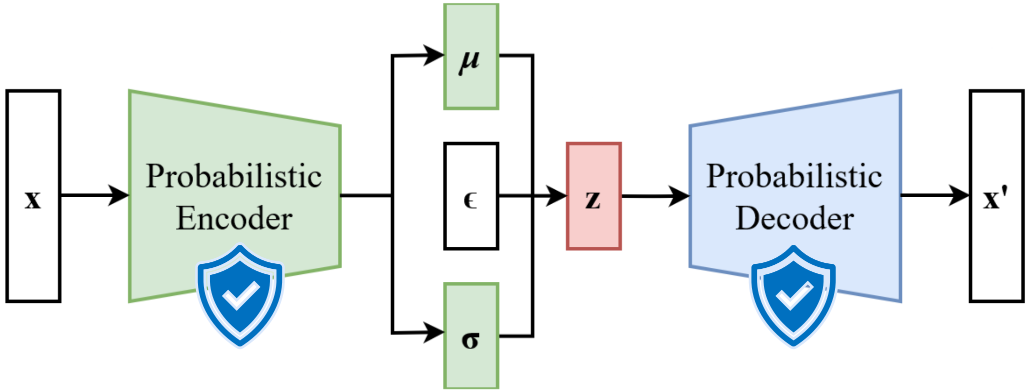


- *One β-Variational Autoencoder model for each class of each dataset per member*
- *Hypothesis: each member has enough hardware capabilities and skills*

[12] Higgins, L. Mattheyet al. "beta-vae: Learning basic visual con-cepts with a constrained variational framework," in 2017 InternationalConference on Learning Representations (ICLR), 2017.

The ingredients

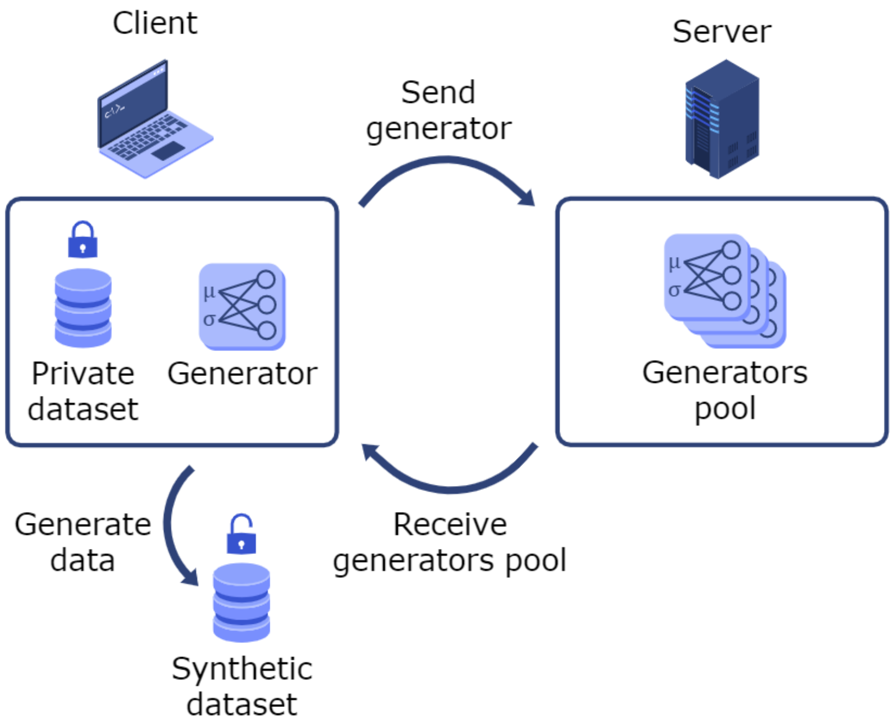




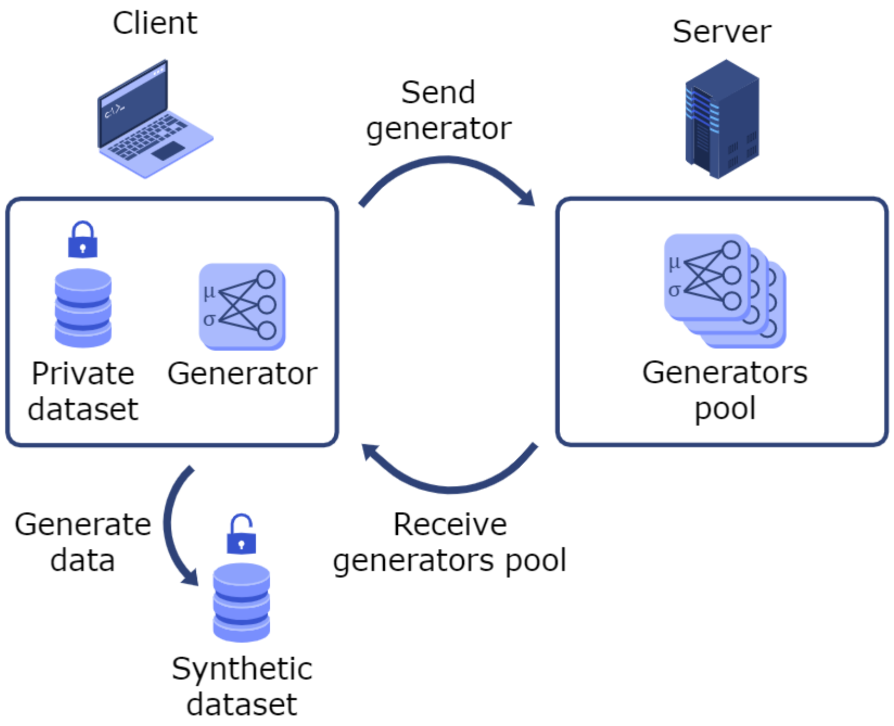
- *One β-Variational Autoencoder model for each class of each dataset per member*
- *Hypothesis: each member has enough hardware capabilities and skills*
- *Every member of the federated party trains all the models with strict (ε, δ)-DP constraints* $(\varepsilon < 2, \delta \leq 10^{-4}, RDP \geq 8)$

[12] Higgins, L. Mattheyet al. "beta-vae: Learning basic visual con-cepts with a constrained variational framework," in 2017 InternationalConference on Learning Representations (ICLR), 2017.

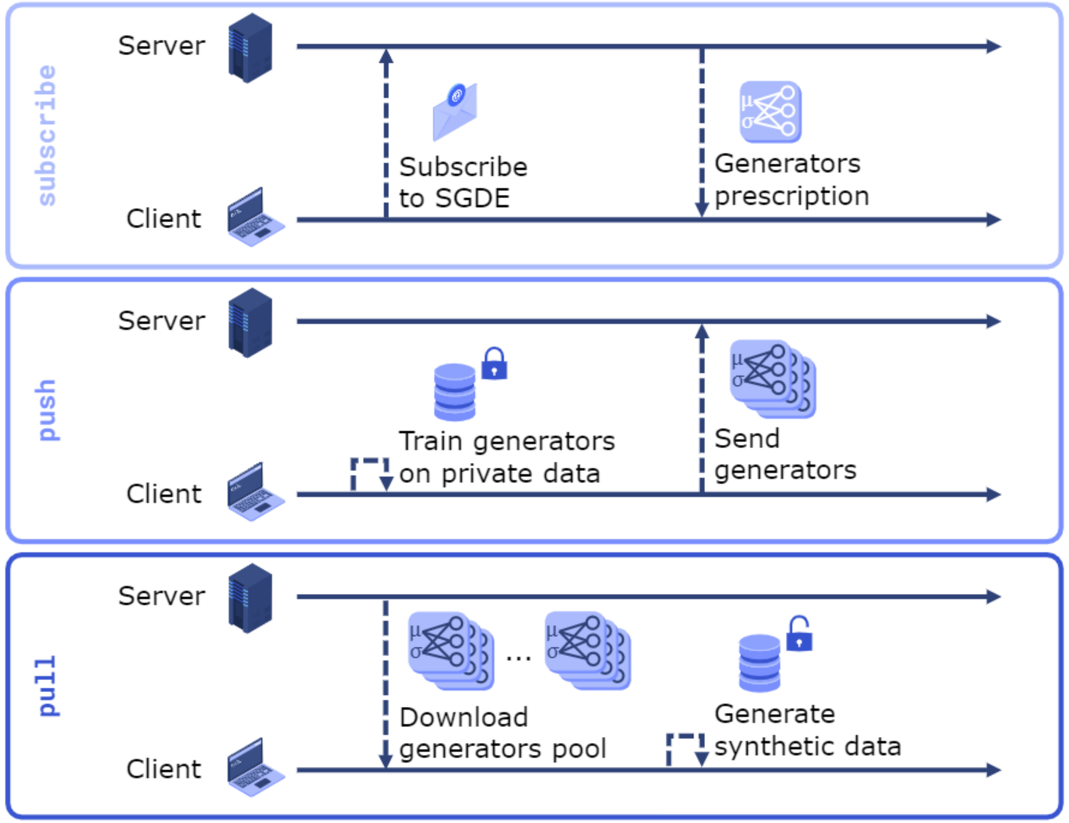
The method


Client
Server
Send
generator
Private
dataset
Generator
Generators
pool
Generate
data
Receive
generators pool
Synthetic
dataset

The method



Client
Server
Send generator
Private dataset
Generator
Generators pool
Generate data
Receive generators pool
Synthetic dataset



subscribe
Server
Client
Subscribe to SGDE
Generators prescription
push
Server
Client
Train generators on private data
Send generators
pull
Server
Client
Download generators pool
Generate synthetic data

The results

Average improvement on local data

Dataset	Accuracy		F1 score		AUC	
	Real	*Synth*	*Real*	*Synth*	*Real*	*Synth*
Titanic	75.67	**80.87**	19.43	**63.37**	75.7	**78.35**
Breast Cancer	89.67	**97.09**	93.37	**97.81**	99.17	**99.27**
Mushrooms	92.93	**93.49**	92.43	**93.14**	96.23	**96.61**
Adult	**80.64**	79.65	49.69	**61.64**	83.30	**83.73**
Wine Quality	93.46	**98.54**	82.98	**97.10**	99.44	**99.49**
MNIST	98.20	**98.72**	98.16	**98.71**	99.02	**99.31**
Fashion MNIST	88.47	**89.30**	88.32	**89.22**	93.87	**94.76**
Avg. Improvement		+2.66		+10.94		+0.68

Average improvement on external data

Dataset	Accuracy		F1 score		AUC	
	Real	*Synth*	*Real*	*Synth*	*Real*	*Synth*
Titanic	71.83	**74.01**	29.70	**56.00**	77.14	**77.43**
Breast Cancer	89.42	**93.02**	92.25	**94.78**	99.60	**99.76**
Mushrooms	92.56	**93.49**	91.92	**93.14**	96.30	**96.61**
Adult	**80.87**	79.00	50.14	**60.21**	84.02	**84.08**
Wine Quality	92.57	**97.79**	82.42	**95.70**	98.63	**98.65**
MNIST	97.76	**98.49**	97.71	**98.49**	99.02	**99.19**
Fashion MNIST	85.97	**88.13**	85.81	**88.04**	92.65	**94.13**
Avg. Improvement		+1.85		+8.06		+0.36

Thank you
for the attention!

References

[1] EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

[2] Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. "Explaining and harnessing adversarial examples." arXiv preprint arXiv:1412.6572 (2014).

[3] N. Bouacida and P. Mohapatra, “Vulnerabilities in federated learning,” IEEE Access, vol. 9, pp. 63 229–63 249, 2021.

[4] Shokri, Reza, et al. "Membership inference attacks against machine learning models." 2017 IEEE symposium on security and privacy (SP). IEEE, 2017.

[5] Liu, Gaoyang, et al. "Socinf: Membership inference attacks on social media health data with machine learning." IEEE Transactions on Computational Social Systems 6.5 (2019): 907-921.

References

[6] Dwork, Cynthia, et al. "Our data, ourselves: Privacy via distributed noise generation." EUROCRYPT. Springer, Berlin, Heidelberg, 2006.

[7] Cynthia, Dwork. "Differential Privacy In Automata, Languages and Programming, Bugliesi Michele, Preneel Bart, Sassone Vladimiro, and Wegener Ingo." (2006): 1-12.

[8]Abadi, Martin, et al. "Deep learning with differential privacy." Proceedings of the 2016 ACM SIGSAC conference on computer and communications security. 2016.

[9] Dwork, Cynthia, and Aaron Roth. "The algorithmic foundations of differential privacy." Found. Trends Theor. Comput. Sci. 9.3-4 (2014): 211-407.

[10] Mironov, Ilya. "Rényi differential privacy." 2017 IEEE 30th computer security foundations symposium (CSF). IEEE, 2017.

References

[11] Krizhevsky, Alex, and Geoffrey Hinton. "Learning multiple layers of features from tiny images." (2009).

[12] Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick,S. Mohamed, and A. Lerchner, “beta-vae: Learning basic visual con-cepts with a constrained variational framework,” in2017 InternationalConference on Learning Representations (ICLR), 2017.