PEARL

Data synthesis via private embeddings and adversarial reconstruction learning

ICLR 2022
arXiv:2106.04590

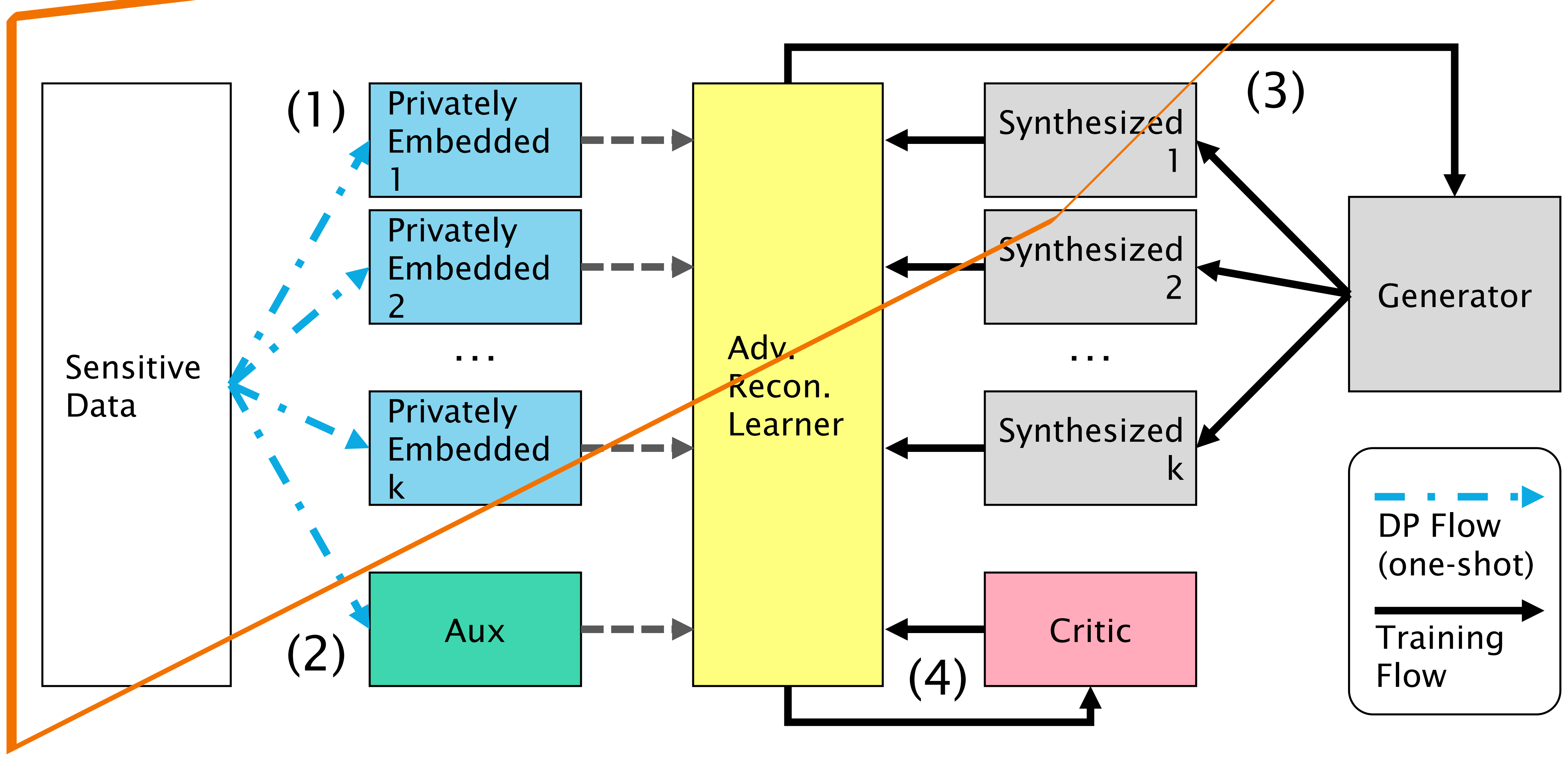
Seng Pei Liew
with Tsubasa Takahashi & Michihiko Ueno

LINE Corporation

LINE

Summary in one slide

- Sharing data among organizations or departments may cause privacy issues (How to mitigate this issue?)
- **Privacy-preserving data synthesis (PPDS)**: we train a *generative model* with differential privacy (rigorous privacy guarantees) and use the model to generate synthetic data for private data sharing purposes
- We propose **PEARL**, a framework to train generative models at practical level of privacy, and overcomes issues encountered in previous works which mainly utilize DP-SGD (to be explained in later parts)



Differentially private data synthesis

Sensitive data

“Fake data” that is private and preserves the characteristics of the real data

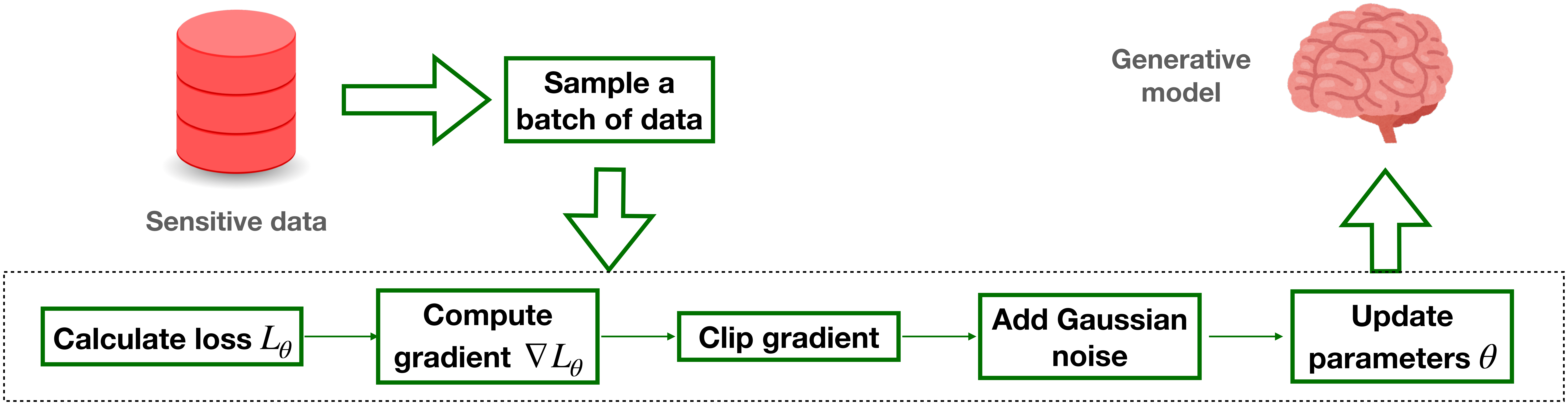
(Data scientist)

Allow arbitrary usage without privacy violation

- **Training ML models**
- **Exploratory data analysis**

Training deep generative models with differential privacy

- The most popular method is differential private stochastic gradient descent (DP-SGD) [ACG+16]
- DP-SGD ensures that each gradient update is private, which in turn guarantees that the network parameters are private



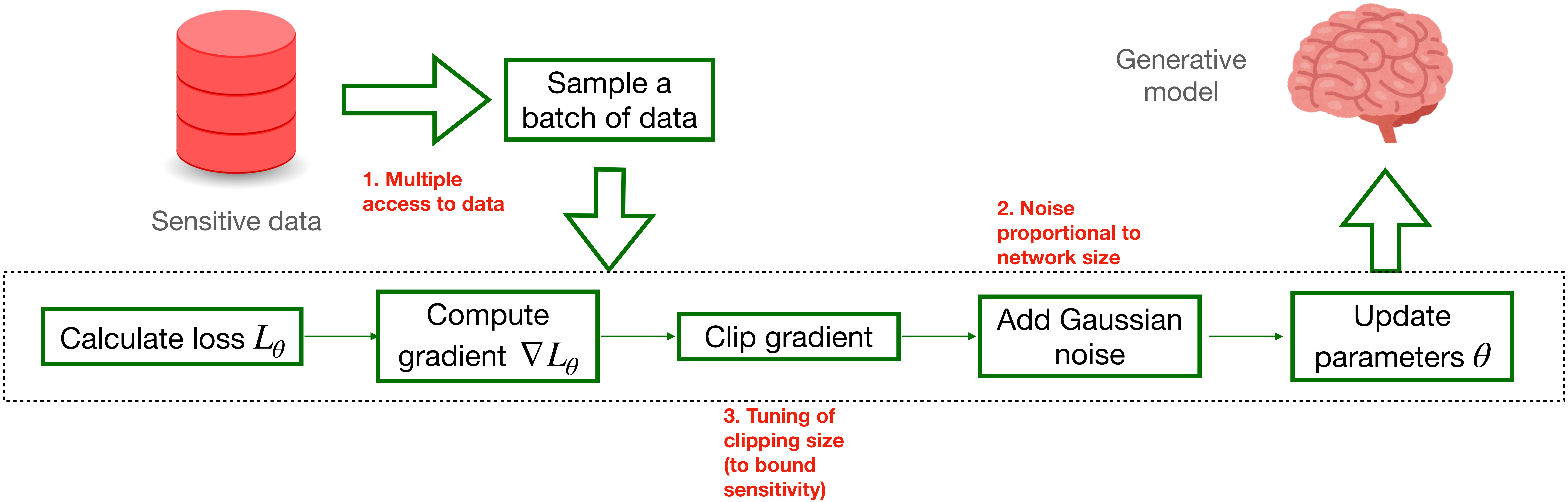
- Accumulate privacy consumption with moments accountant.

General shortcomings of DP-SGD

1. Training steps are limited. Each access of data reduces the guarantees of privacy.

2. Network size is limited. Large neural networks lead to too much noises added to the gradient updates.

3. Extensive hyperparameter (clipping size) tunings are required.



Proposal: PEARL

Private **E**mbeddings and **A**dversarial **R**econstruction **L**earning (arXiv: 2106.04590)

1. Project sensitive data to low-dimensional embeddings and add Gaussian noises to make the embeddings differentially private

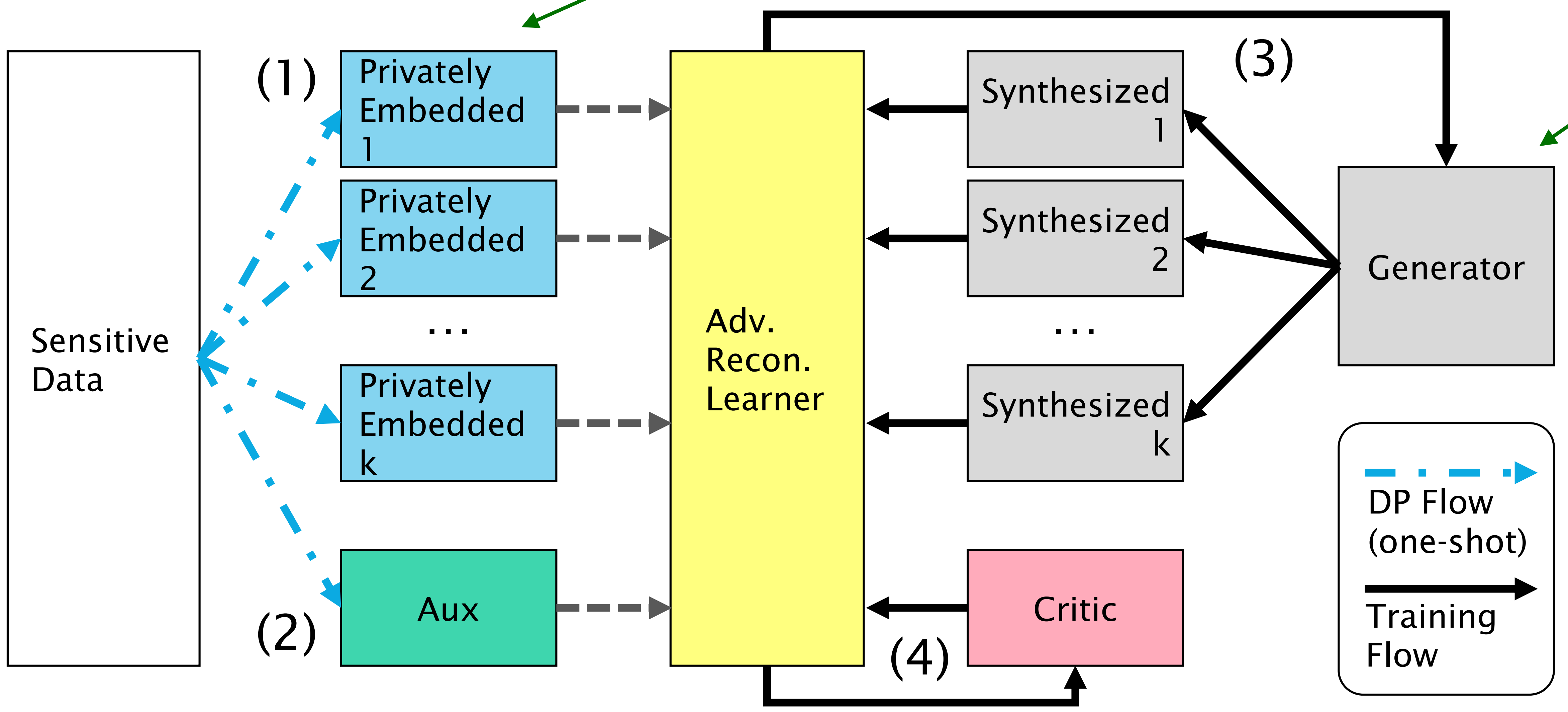
2. Obtain auxiliary information useful for training in a differential private manner

3. Train a generator by minimizing the embedding distance

4. Train with an adversarial objective to improve the performance

One-shot data access

No noise added to the gradients. No clipping required.



Realization of PEARL

Characteristic Function

- Let $\mathbf{x}$ be a random variable with probability distribution $\mathbb{P}$, the corresponding characteristic function is

$$\Phi_{\mathbb{P}}(t) = \mathbb{E}_{\mathbf{x}\sim\mathbb{P}}[e^{it\cdot\mathbf{x}}] = \int_{\mathbb{R}^d} e^{it\cdot\mathbf{x}} d\mathbb{P} \simeq \sum_{\mathbf{x}} e^{it\cdot\mathbf{x}}$$ **(empirical CF)**

- This mathematical operation is equivalent to *Fourier transformation* from the signal processing point of view. t is frequency.

- Also define Characteristic function distance between two distributions:

$$C^2(\mathbb{P}, \mathbb{Q}) = \int |\Phi_{\mathbb{P}}(t) - \Phi_{\mathbb{Q}}(t)|^2 \omega(t) dt$$

- It can be shown that with appropriately defined density $\omega(t)$, $C(\mathbb{P}, \mathbb{Q}) = 0 \iff \mathbb{P} = \mathbb{Q}$

Realization of PEARL

- The following minimax optimization is proposed:

$$\boxed{\inf_{\theta \in \Theta} \sup_{\omega \in \Omega} \sum_{i=1}^{k} \frac{\omega(t_i)}{\omega_0(t_i)} \left| \widetilde{\Phi}_{\mathbb{P}}(t_i) - \widehat{\Phi}_{\mathbb{Q}}(t_i) \right|^2}$$

- Additionally, we are able to show that the above optimization has the following theoretical properties:

    1. Continuity and differentiability (allows generator to be trained via gradient descent)

    2. Weak convergence (good for training GAN-like models [ACB'17])

    3. Consistency at infinite sampling limit (ensures the maximization procedure is consistent asymptotically)

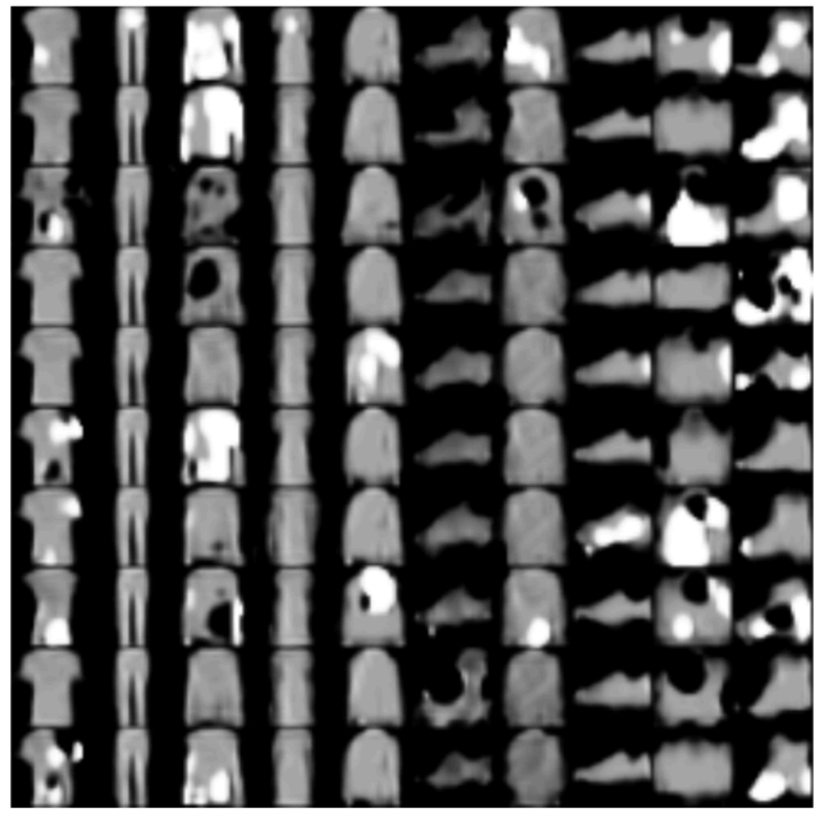
Generated image data

(a) $\epsilon = \infty$

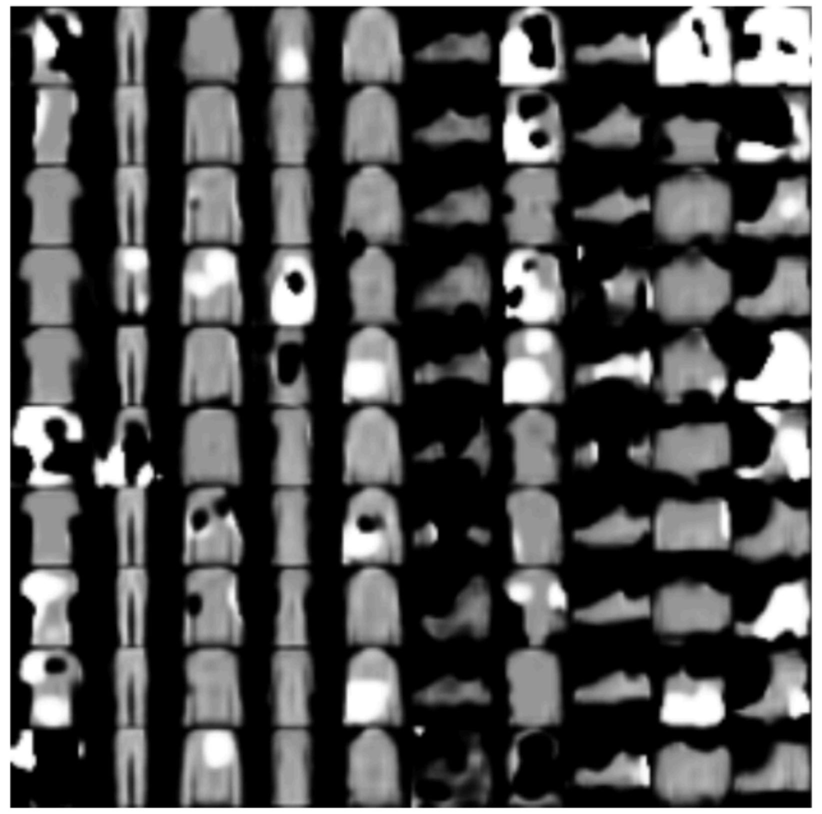
(b) $\epsilon = 10$

(c) $\epsilon = 1$

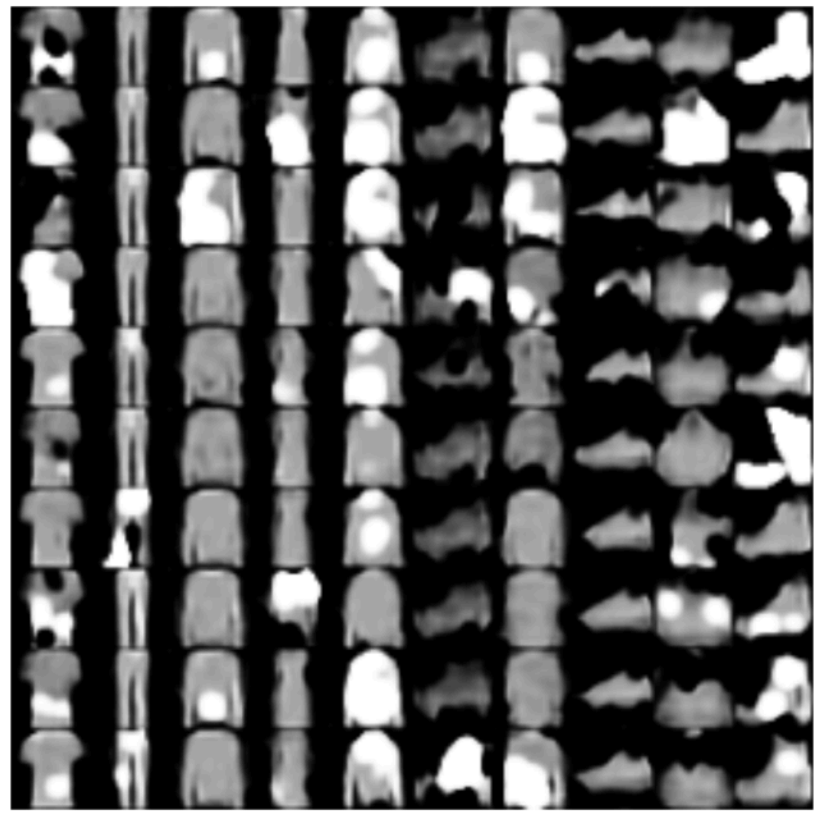
(d) $\epsilon = 0.2$



(a) $\epsilon = \infty$



(b) $\epsilon = 10$



(c) $\epsilon = 1$

(d) $\epsilon = 0.2$

- PEARL's quality is low at non-private ($\epsilon = \infty$) limit, but the quality doesn't change much as ϵ decreases (except at extreme value)

Generated image data

Datasets	Metrics	DP-MERF	Ours (Min only)	Ours (Minimax)
MNIST	FID	49.9 ± 0.22	3.79 ± 0.06	$\mathbf{3.52 \pm 0.06}$
	KID $(\times 10^3)$	148 ± 46.2	77.8 ± 9.88	$\mathbf{70.5 \pm 10.3}$
Fashion-MNIST	FID	37.0 ± 0.15	1.99 ± 0.04	$\mathbf{1.92 \pm 0.04}$
	KID $(\times 10^3)$	1220 ± 36.1	$\mathbf{24.0 \pm 6.90}$	26.9 ± 6.80

Table 1: FID and KID (lower is better) on image datasets at $(\epsilon, \delta) = (1, 10^{-5})$.

- Evaluating with metrics commonly used for GANs

Results on tabular data

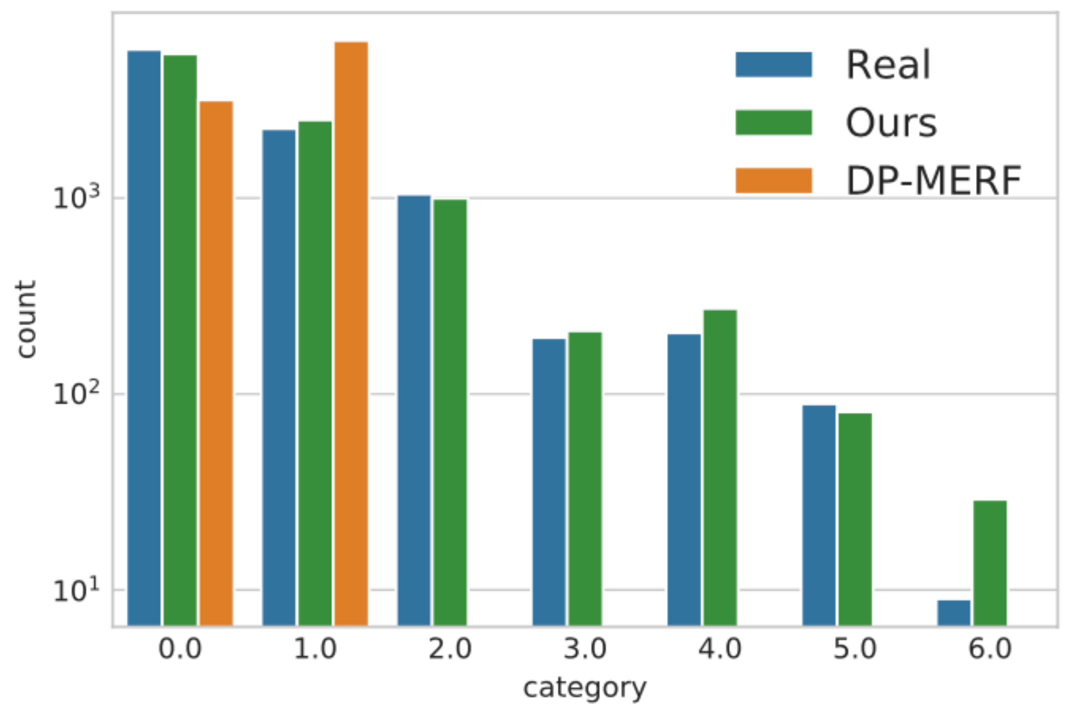


Figure 4: Plot of histogram for the “marital-status” attribute of the Adult dataset. Evaluation is performed at $(\epsilon, \delta) = (1, 10^{-5})$.

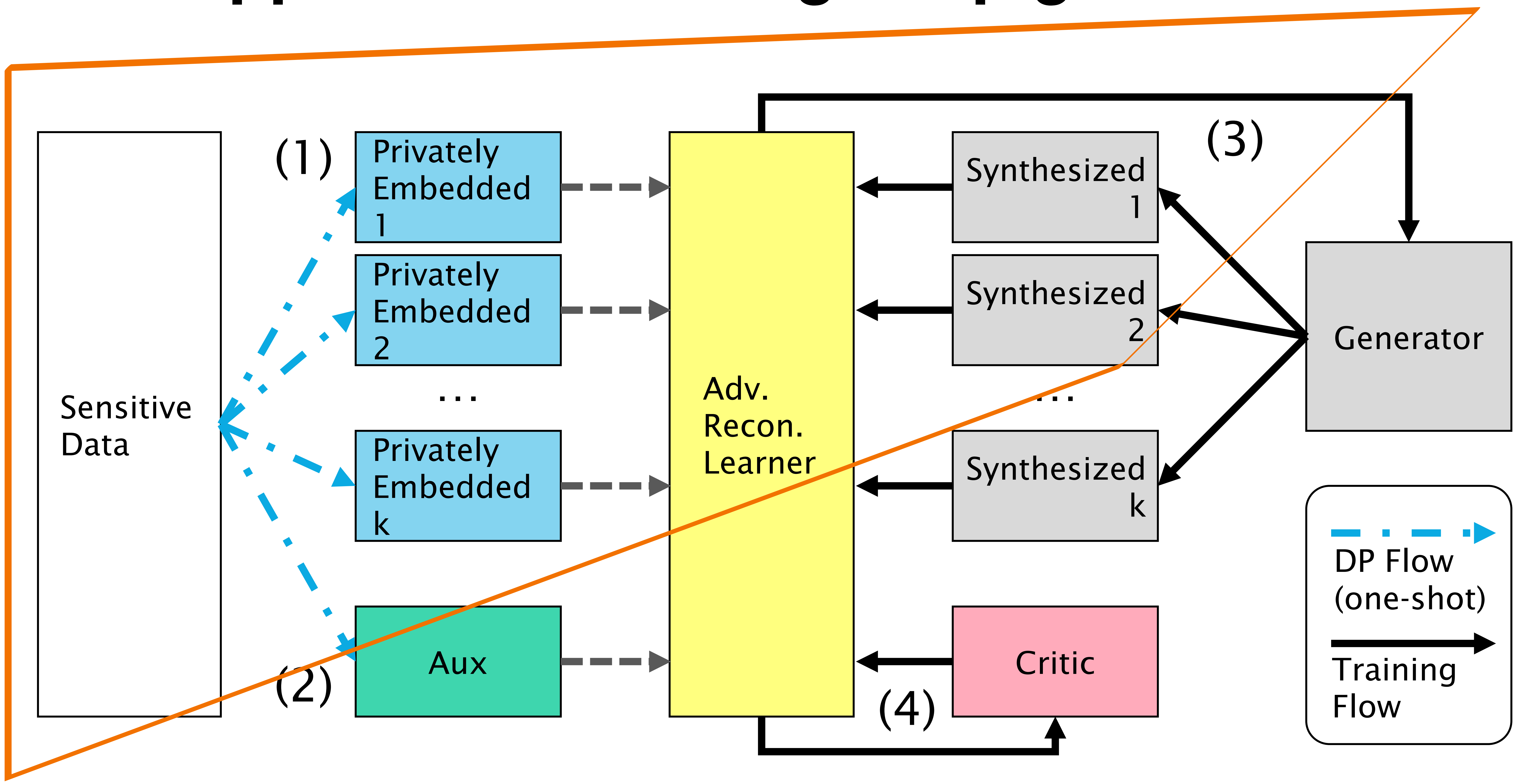
Data	Metrics	Average
Real data	ROC	0.765 ± 0.047
	PRC	0.654 ± 0.050
DP-MERF	ROC	0.641 ± 0.044
	PRC	0.536 ± 0.034
Ours	ROC	$\mathbf{0.721 \pm 0.035}$
	PRC	$\mathbf{0.618 \pm 0.033}$

Table 2: Average ROC and PRC scores for the Adult dataset evaluated at $(\epsilon, \delta) = (1, 10^{-5})$.

- We also generate synthetic *Adult* data. The frequency histogram is shown in the left (compared with another SOTA method), which can capture the pattern of the distribution well.
- We use the synthetic data to train ML models for classifying real data. The result on the right also show that PEARL outperforms the SOTA method.

Wrap-up

PEARL: a new approach of training deep generative models



- Training practical models at reasonable privacy levels while avoiding difficulties of DP-SGD.