Reliable and Trustworthy Artificial Intelligence

Lecture 9: AI Regulations and Synthetic Data

Martin Vechev, Mislav Balunović

ETH Zurich

Fall 2022

Motivation - sensitive applications of AI

AI is increasingly used for decision making in sensitive industries:

Banking

Insurance

Healthcare

Human Resources

UK watchdogs to clamp down on banks using discriminatory AI in loan applications

Will Paige Feb 15, 2022, 4:48 PM

Dutch scandal serves as a warning for Europe over risks of using algorithms

The Dutch tax authority ruined thousands of lives after using an algorithm to spot suspected benefits fraud – and critics say there is little stopping it from happening again.

NYC Law Restricting Use of AI in Hiring Takes Effect in January: Are You Ready?

Thursday, September 1, 2022

Report: AI Company Leaks Over 2.5M Medical Records

The leaked data relates to car accidents and includes names, insurance records, medical diagnosis notes, and payment records.

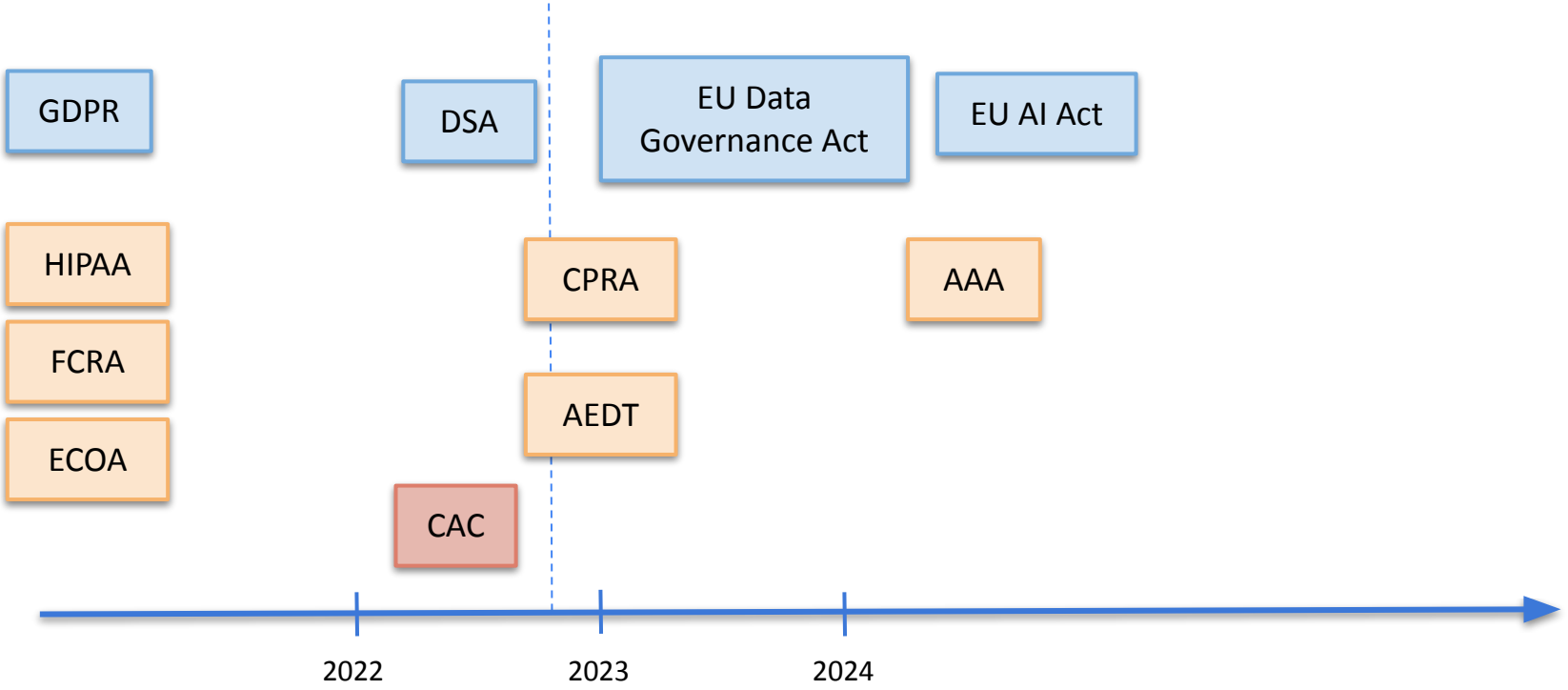
Geographic overview of AI regulations

GDPR,
EU AI Act

AAA,
AI Bill of Rights,
CPRA

CAC Recommendation
Algorithms Regulation

Timeline of Data-related and AI Regulations



Fines for violations of existing AI regulations (e.g., GDPR)

France: CNIL fines Clearview AI €20 million over facial recognition technology

CNIL outlined that Clearview AI's facial recognition software is based on the systematic and widespread collection of images containing faces without consent.

China Imposes $1.2 Billion Fine for Data Violations

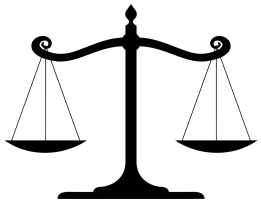
China fined its largest ride-hailing company for violating data protection laws. They illegally collected large amount of data on passengers.

Tax Administration fined for discriminatory and unlawful data processing

The Dutch Data Protection Authority (DPA) has imposed a €2.75 million fine on the Dutch Tax Administration for violating data minimization principle when training AI for fraud detection.

Fairness

Fairness regulations state that AI is not allowed to discriminate.



Fairness is captured in many existing (FCRA, ECOA, GDPR) and upcoming (AEDT, EU AI Act) regulations.

Why can AI discriminate? Often the culprit is data, e.g. biased data or data which does not contain enough minority samples, but it can also be because of the training algorithm.

Older regulations (created before AI but apply to AI as well)

There are older regulations for denying employment, housing, credit, insurance. These regulations were created when such decision-making was done by humans, but they still apply now when AI makes decisions.

Section 5 of the FTC Act (1914). Prohibits unfair or deceptive practices such as sale or use of racially biased algorithms.

Fair Credit Reporting Act (1970). Applies when an algorithm is used to deny people employment, housing, credit, insurance, or other benefits.

Equal Credit Opportunity Act (1974). Illegal for a company to use a biased algorithm that results in credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age.

NYC regulation for automated employment decision tools (AEDT)

Enters into practice on January 1, 2023

Prohibits employers from using AEDT unless such tool has been subject to a bias audit within one year of the use of the tool

Impact ratio concrete metric: compute impact ratio for each category (e.g. Black female) by dividing selection rate of the category by the selection rate of the most selected category. Employers need to compute this metric and make it available.

Aspects of the fairness problem

Sensitive attribute: Each regulation and each application (e.g., insurance) defines their own sensitive attribute. For example, in case of EU regulation and insurance application, sensitive attribute is gender.

Fairness metric: Regulation might define a specific fairness metric or it can broadly state that AI should not discriminate. For example, FTC (Federal Trade Commission) defines discrimination if impact ratio is below 80% (so-called four-fifths rule). Metric is actually used by some Swiss insurance companies.

How do we train provably fair AI models? Next set of lectures!

Explainability

ECOA: Creditors are required to notify applicants who are denied credit with specific reasons for the detail.

GDPR: The data subject should have the right to obtain an explanation of the decision reached.

There are several criticisms of this requirement:

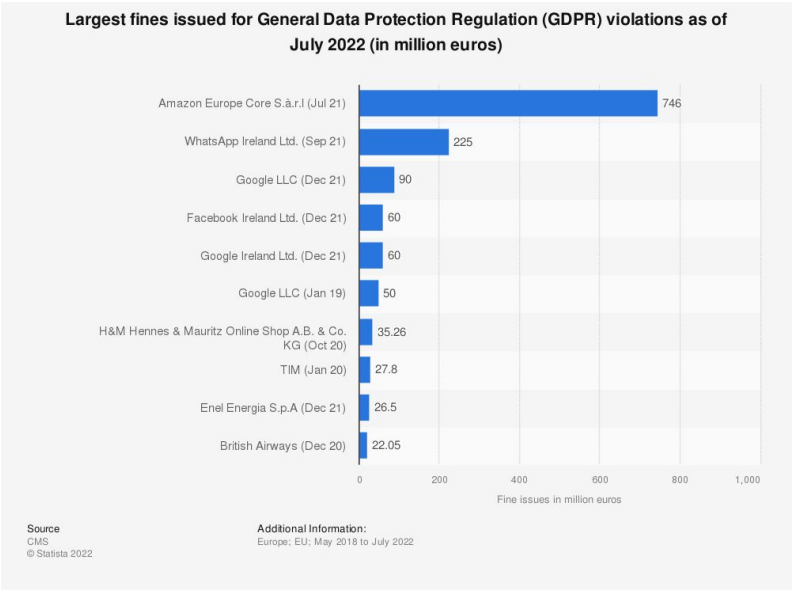
- Human decisions themselves are often not explainable
- Output of deep neural networks is not explainable (though there is active research) so it might hurt innovation by restricting the usage of advanced models

GDPR - General Data Protection Regulation

GDPR contains rules relating to the protection of people with regard to the processing and movement of personal data in EU.

GDPR was adopted in 2016 and inspired many similar regulations (e.g. California Privacy Rights Act - CPRA).

So far issued fines in total of more than **2 billion €**



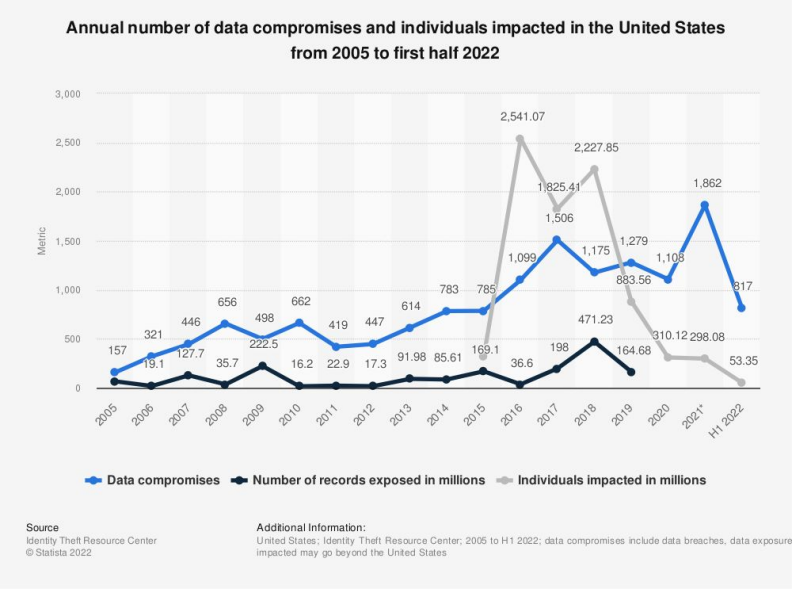
Violating GDPR: Risks of sharing data without privacy protection

ML systems are vulnerable to software bugs as any other system.

Even libraries such as NumPy and Pickle have shown to be vulnerable to remote code execution if not used safely. See: https://huggingface.co/docs/hub/security-pickle

Sharing data with external parties is always risky!

Hackers do not need to perform sophisticated attacks (e.g. membership inference attacks) if they can just exploit some vulnerability and get the raw data



*attacks on data leaks are increasing (blue line)

GDPR and AI

While GDPR was not created with AI in mind, it applies to all automated decision making systems

There is a tension between AI which requires more and more data, and GDPR that is trying to limit the collection and sharing of data

Collecting and sharing data increases risk of data leakage

GDPR - Defines different data types

Personal data - Any information relating to an identified or identifiable natural person. For example, medical records containing name of the person.

Pseudo-anonymized data - Personal data that can no longer be attributed to a data subject without the use of additional information. For example, names are replaced by reference codes that are stored in a separate database.

Anonymized data - Personal data rendered anonymous so that the data subject is not or no longer identifiable. For example, names are replaced by completely random ones. This is the only data type out of scope of GDPR.

GDPR - Identifies privacy risks

Singling out - Locating individual's record in the dataset. For example, given info about a person, find their record in a leaked database of medical records.

Linkability - Linking two records of the same individual. For example, link the address in leaked database of food delivery to medical records to find health status of a person.

Inference - Estimating personal data from the attributes in the record. For example, estimate the health status from a leaked database of fitness activity.

How do we formalize GDPR privacy risks (e.g. Cohen & Nissim 2020 do it for 'singling out')?

https://arxiv.org/abs/1904.06009

GDPR related: promising paradigms to protect data but still work-in-progress

Some approaches help to avoid collecting and storing user data:

- Federated Learning (discussed already in prior lectures)
- Trusted Execution Environment (e.g., SGX extensions)
- Homomorphic encryption, Secure MPC
- Synthetic data (closer look in 2nd part of the lecture)

What if we need to collect the data?

GDPR - Data Minimization

The data minimization principle is expressed in Article 5(1)(c) of the GDPR and Article 4(1)(c) of Regulation (EU) 2018/1725, which provide that personal data must be "**adequate**, **relevant** and **limited to what is necessary** in relation to the purposes for which they are processed".

Questions:

How do we measure the amount of collected data?

Are all data points needed to achieve good accuracy?

Are all collected users' features needed to achieve good accuracy?

Open problem: Formalizing and achieving data minimization

GDPR - Data Minimization Example: Insurance (E)ligibility

Age	Nat	Zip	Salary	Smoke	Job	E
37	CH	8020	85K	True	Pharma	0
26	US	1000	60K	False	Engineer	1
52	CH	7050	100K	True	IT	1
24	DE	2055	48K	False	Driver	0
62	IT	1505	65K	False	Nurse	0

Age	Nat	Salary	Smoke	Job	E
[30, 40]	CH	[80K, 90K]	True	Pharma	0
[20, 30]	US	[50K, 60K]	False	Engineer	1
[50, 55]	CH	[90K,100K]	True	IT	1
[20, 30]	EU	[40K, 50K]	False	Driver	0
[60, 65]	EU	[60K, 70K]	False	Nurse	0

Input dataset

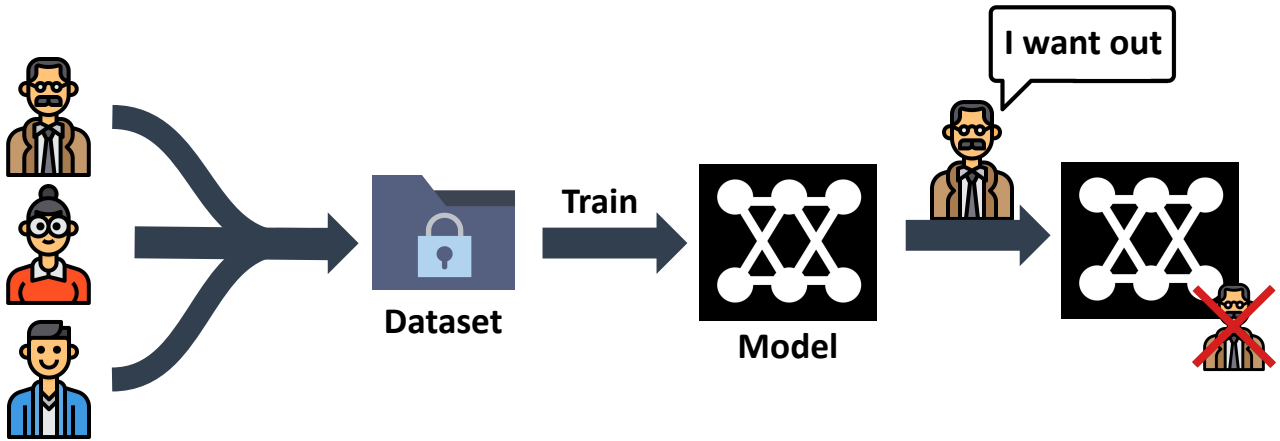
Output dataset

Key challenge: Remove features that are not necessary to predict insurance eligibility and only keep what is necessary to solve the task

GDPR - Unlearning

Motivation:

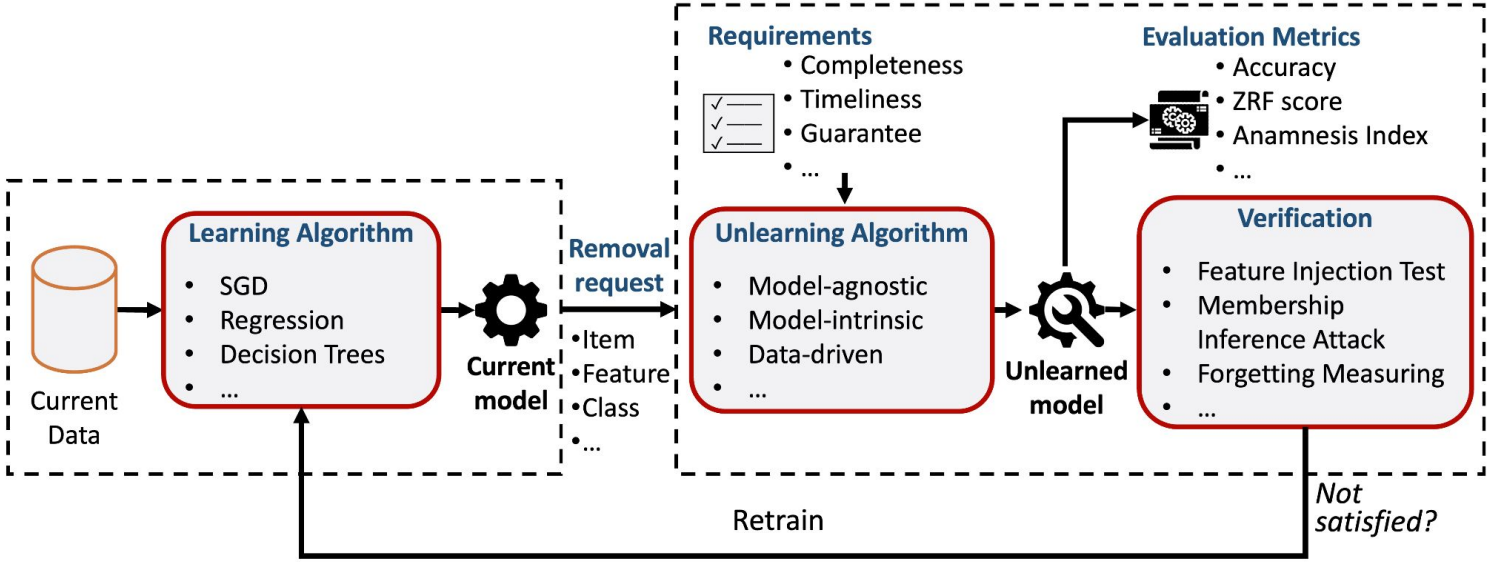
- Right to be forgotten (Article 17 of GDPR) - Users can withdraw their data consent
- Often user consent has a time limit



Goal: Users should be able to opt out of participation

Machine unlearning, SP 2021, Bourtoule et. al https://arxiv.org/abs/1912.03817

GDPR - Unlearning: Technical challenges



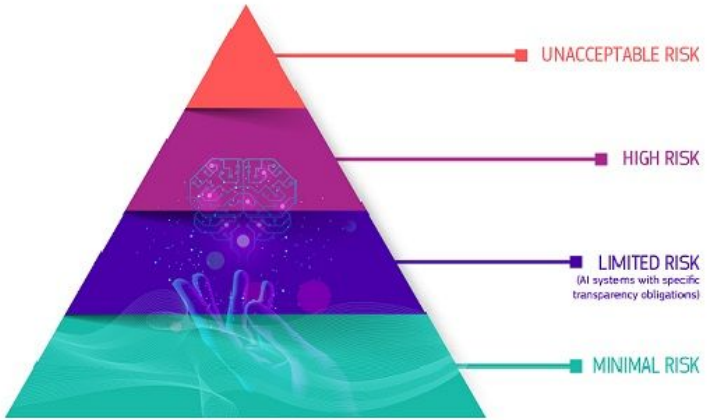
- **Accuracy:** it should not degrade
- **Consistency:** weights should be consistent with weights of retrained model
- **Timeliness:** unlearning should be faster than simply retraining the model
- **Guarantees:** prove that some sample is indeed unlearned

Key challenges

EU AI Act

Expected to become active in 2024, still changing. Does not really discuss privacy (main focus of GDPR). The regulatory framework defines 4 levels of risk of AI usage:

- Unacceptable risk
- High risk
- Limited risk
- Minimal or no risk



EU AI Act: Unacceptable risk

Social scoring - AI systems classify the trustworthiness of natural persons based on their social behaviour in multiple contexts. For example, such social score derived from shopping data could be used to restrict travel.

Distorting human behavior - AI systems deploy subliminal components that exploit vulnerabilities of children and people due to their age, physical or mental incapacities. For example, AI could try to detect very old people and show them ads for expensive medicine.

Real-time biometric information for law enforcement - For example, city could be covered by surveillance cameras and movement of citizens tracked by AI face recognition.

EU AI Act - High risk

AI used for:

Critical infrastructure - e.g. safety components of water supply

Access to education - e.g. accepting a person to university

Employment - e.g. making promotion decisions

Public and private services - e.g. credit score evaluation

Migration - e.g. AI-based polygraph at border control

Administration of justice - e.g. assisting judges in criminal cases

EU AI Act - Limited and minimal risk

AI used for:

AI chatbots - e.g. customer support

AI enabled video games - e.g. AI opponents in the video games

AI spam filters - e.g. for mail

Most of the other systems - e.g. in manufacturing

EU AI Act - Requirements for high risk AI

Risk management system - A continuous iterative process run throughout the entire lifecycle of a high-risk AI system

Data governance - Training, validation, and test set should meet quality criteria

Technical documentation - Documentation should demonstrate that AI system complies with requirements

Record-keeping - They should automatically record events while the AI system is operating

Transparency - Operation of the system should be sufficiently transparent to enable users to interpret system output

Human oversight - Systems should be effectively overseen by natural persons during their use

Accuracy, robustness, cybersecurity - Systems should achieve appropriate levels of accuracy, robustness, and cybersecurity during their lifecycle

(U.S.) AI Bill of Rights

Not a regulation, but a foundation for upcoming regulations in US, e.g. Algorithmic Accountability Act (AAA)

AI Bill of Rights states that people should:

- Be protected from safe and ineffective systems
- Not face discrimination by algorithms and systems
- Be protected from abusive data practices
- Know that automated system is used and how it impacts the outcome
- Have an option to opt out and have access to human alternative where appropriate

Digital Services Act (not explicit about AI but affects AI), more towards regulating large platforms

Entered October 19, 2022

Obligate platforms which have num. of users at least 10% of EU to have their products assessed by independent party (including algorithms) for societal risks (e.g. Facebook)

Transparency reports required for automated systems used for e.g. moderation (e.g., removing YouTube videos may favor one political party over another)

Limits on advertising platforms that may manipulate users (e.g. make sure that GoogleAds does not manipulate users to buy some products).

CAC Recommendation Algorithms Regulation: Entered March 2022

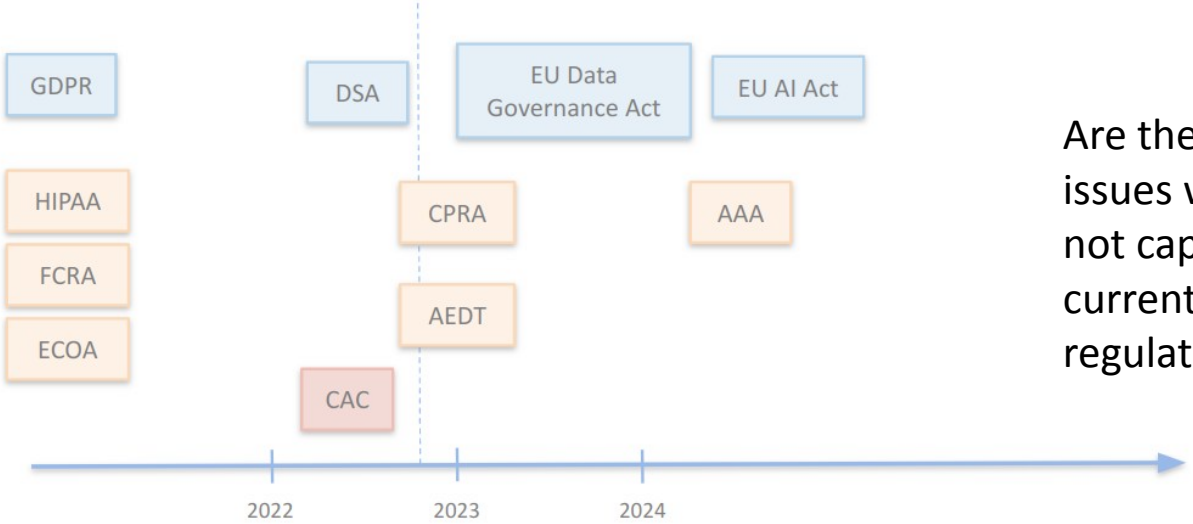
https://digichina.stanford.edu/work/experts-examine-chinas-pioneering-draft-algorithm-regulations/

Ground breaking regulation for recommendation algorithms (e.g. TikTok)

Requires that users should:

- Have right for explanation
- Be able to edit tags used for recommendation by themselves
- Not be subject to differential treatment based on their preferences (e.g., Differential treatment - users see different price based on their past consumer activity (extracted from collected data)).
- Not be subject to algorithms that encourage addictive behavior

Timeline of AI Regulations: What might come next?



Are there emerging issues with AI that are not captured by the current (or planned) regulations?

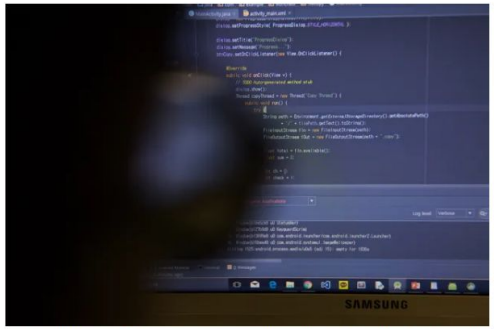
Emerging issues: Copyright + AI

Most advanced AI is trained with large amount of data, which often contains copyrighted material.

For example, Github Copilot may be trained on copyrighted code and Stable Diffusion on copyrighted images.

ARTIFICIAL INTELLIGENCE / TECH / LAW

The lawsuit that could rewrite the rules of AI copyright



The key question in the lawsuit is whether open-source code can be reproduced by AI without attached licenses. Credit: Getty Images

/ Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI. But the suit could have a huge impact on the wider world of artificial intelligence.

By JAMES VINCENT
Nov 8, 2022, 5:09 PM GMT+1 | 8 Comments / 8 New

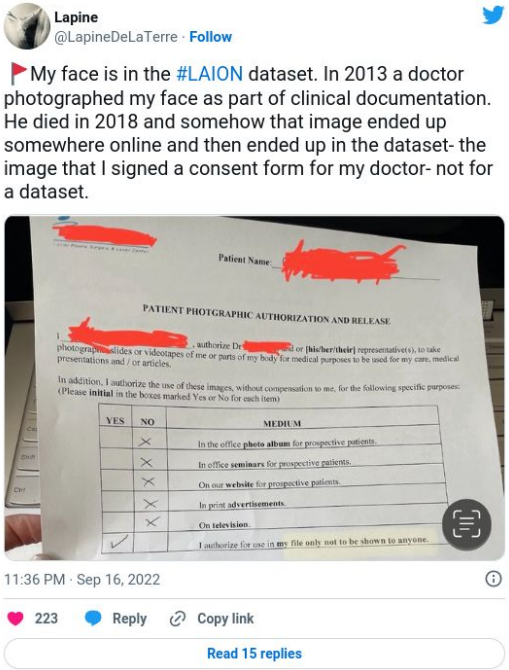
AI Creating 'Art' Is An Ethical And Copyright Nightmare

'A.I. Should Exclude Living Artists From Its Database,' Says One Painter Whose Works Were Used to Fuel Image Generators

Emerging issues: AI generators trained on sensitive data

Large-scale AI is trained on basically everything posted on the internet (no regulation yet stating what is allowed)

These models can be used to generate sensitive content (e.g. deepfakes)



Open problem: Can we somehow perform unlearning on such big models?

Emerging issues: Large scale fake content

Field is moving very fast (example from last week)

Galactica is an LLM for science, trained on 48 million examples of scientific articles, etc.

It turned out Galactica often generates fake papers and articles

Why Meta's latest large language model survived only three days online

Closer look: Synthetic data

Recall from earlier that there are some approaches that aim to avoid sharing sensitive user data

In this part of the lecture, we take a closer look at one of the methods for generating synthetic data with differential privacy guarantees

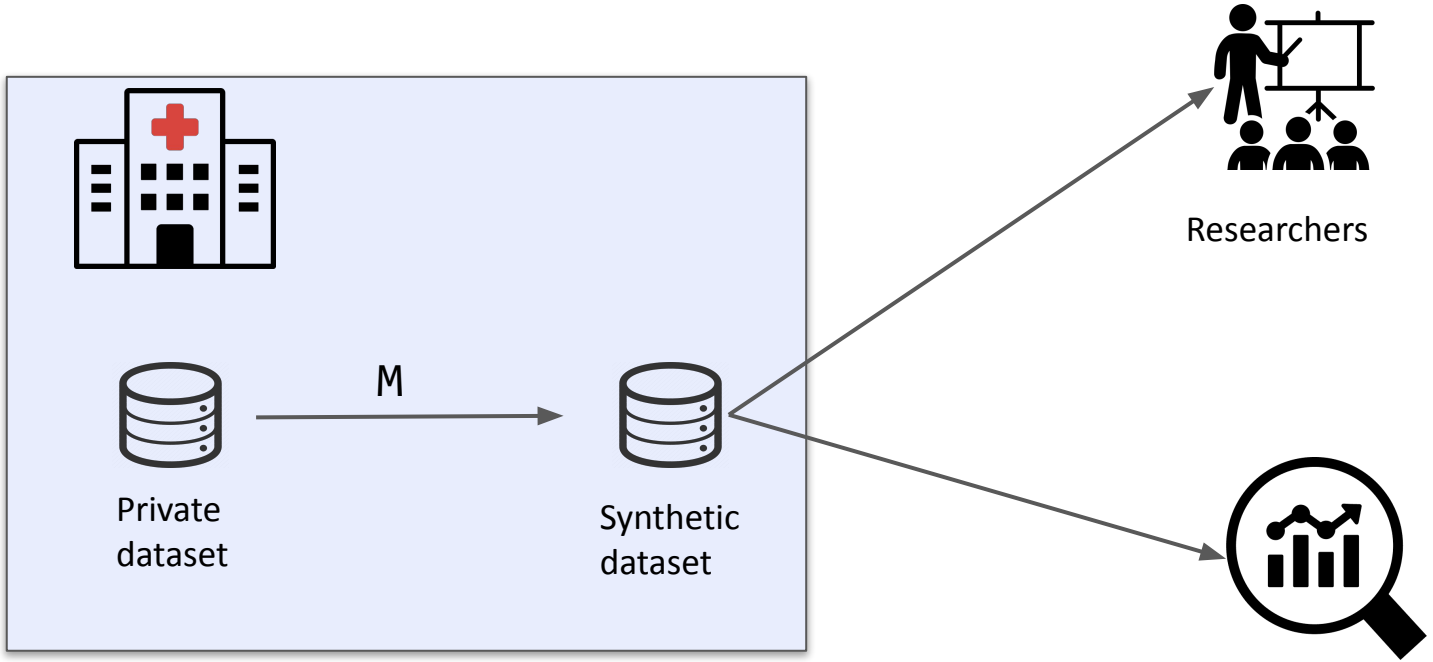
Motivation for synthetic data

Organizations have large amount of data that is sensitive and cannot easily be shared with other parties or between different teams inside the company due to various regulations (e.g. GDPR).

The idea is to generate a new dataset which has the same statistical properties as the original one, but has provable differential privacy guarantees.

Now, instead of sharing the original data, a company can generate synthetic data similar to it and safely share it as it cannot be linked to any individual.

Application of synthetic data



We need to ensure that M is (ε, δ)-DP!

Synthetic data - Example

Age	Nat	Salary	Smoke	E
30	CH	80K	True	0
20	US	50K	False	1
50	CH	90K	True	1
20	EU	40K	False	0
60	EU	60K	False	0

M

Age	Nat	Salary	Smoke	E
20	EU	50K	False	0
45	CH	100K	True	1
60	CH	50K	False	0
30	EU	80K	True	0
20	US	60K	False	1

Input dataset

Output dataset

Key challenge: Generate synthetic data that has similar statistical properties as the original data, but cannot be linked to the real individuals

Synthetic data in industry

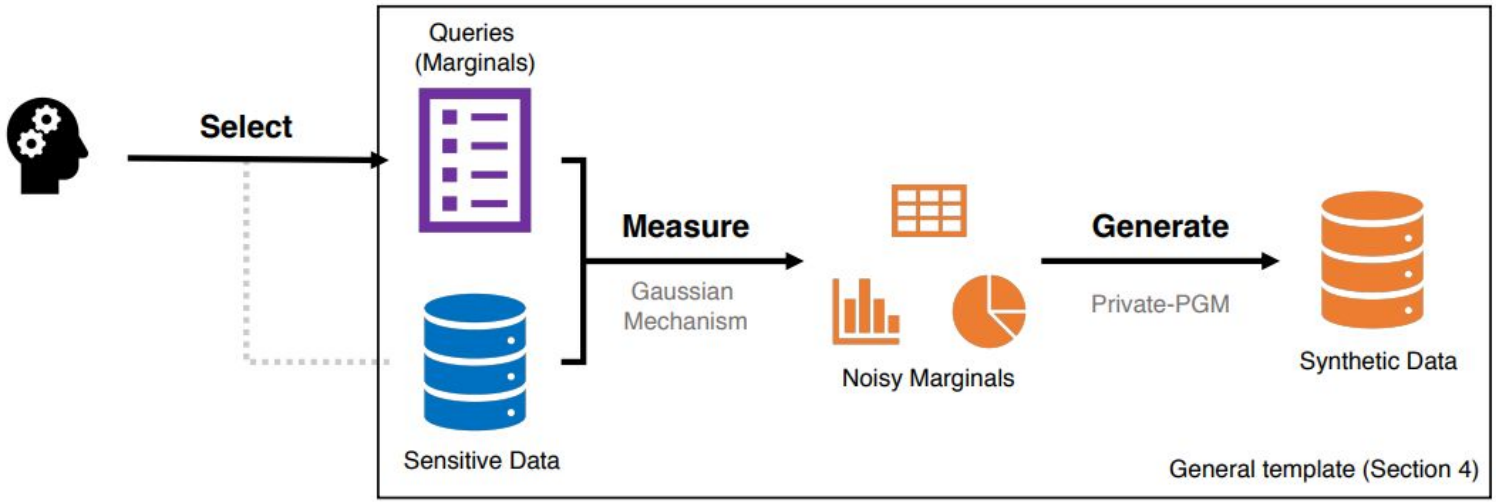
There is a lot of interest in the synthetic data space on the industry side

According to a widely referenced Gartner study, 60% of all data used in the development of AI will be synthetic rather than real by 2024.

Gretel AI raises $50M for a platform that lets engineers build and use synthetic data sets to ensure the privacy of their actual data

Synthetic Data Is About To Transform Artificial Intelligence

Select-Measure-Generate principle: Preview



McKenna, R., Miklau, G., & Sheldon, D. (2021). Winning the nist contest: A scalable and general approach to differentially private synthetic data. *arXiv preprint arXiv:2108.04978*.

Select-Measure-Generate principle

1. Select marginal queries we want to measure
2. Measure marginal queries using differential privacy
3. Generate synthetic data

We first explain the procedure when we do not care about privacy, and then explain how to modify the procedure to be differentially private.

Marginals

Definition 1 (Marginal). *Let $C \subseteq \mathcal{A}$ be a subset of attributes, $\Omega_C = \prod_{i \in C} \Omega_i$, and $n_C = |\Omega_C|$. The marginal on C is a vector $\mu \in \mathbb{R}^{n_C}$, indexed by domain elements $t \in \Omega_C$, such that each entry is a count, i.e., $\mu_t = \sum_{x \in D} \mathbb{1}[x_C = t]$. We let $M_C : \mathcal{D} \to \mathbb{R}^{n_C}$ denote the function that computes the marginal on C, i.e., $\mu = M_C(D)$.*

Definition 6 (Sensitivity). *Let $f : \mathcal{D} \to \mathbb{R}^p$ be a vector-valued function of the input data. The L_2 sensitivity of f is $\Delta_f = \max_{D \sim D'} \|f(D) - f(D')\|_2$.*

Marginal function has sensitivity of 1 because adding a row in a dataset can only change single element of the vector.

Example of a dataset

Smoke	Nat	Kids
True	CH	1
True	Other	0
False	CH	2
True	EU	2
True	CH	1
False	EU	2
False	CH	0
True	Other	1

We have sensitive medical dataset consisting of 3 different features:

Smoke - whether person is a smoker or not
Nationality - nationality of a person
Kids - number of kids the person has

We are interested in generating synthetic dataset based on this one.

Marginals: Example of computation

Smoke	Nat	Kids
True	CH	1
True	Other	0
False	CH	2
True	EU	2
True	CH	1
False	EU	2
False	CH	0
True	Other	1

Original dataset

Smoke	M_C
True	5
False	3

Nat	M_C
CH	4
EU	2
Other	2

Kids	M_C
0	2
1	3
2	3

1-way marginals

Nat/ Smoke	M_C		
	CH	**EU**	**Other**
True	2	1	2
False	2	1	0

Kids/ Smoke	M_C		
	0	**1**	**2**
True	1	3	1
False	1	0	2

Kids/ Nat	M_C		
	0	**1**	**2**
CH	1	2	1
EU	0	0	2
Other	1	1	0

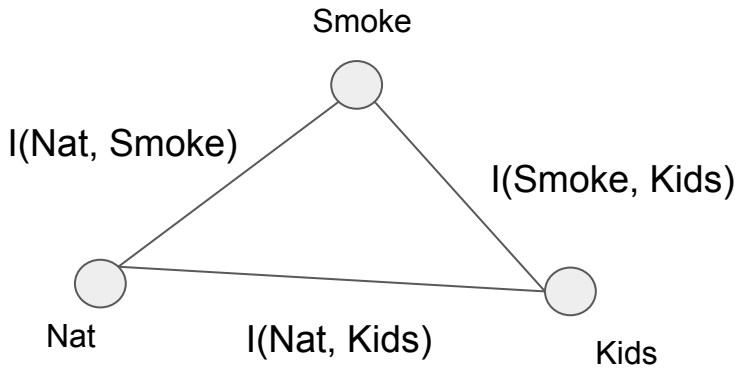
2-way marginals

Selection phase

Mutual information measures mutual dependence between two variables

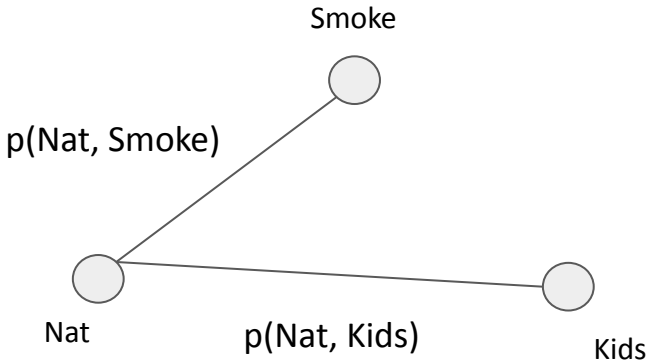
$$I(X,Y) = \sum_x \sum_y \frac{p(X=x, Y=y)}{p(X=x)p(Y=y)}$$

Chow-Liu algorithm: Assign weight to each edge equal to the mutual information between the two variables. Then, compute maximum spanning tree of the resulting graph. It can be shown that this is the optimal second-order approximation.



Inference in a graphical model example

We can generally perform inference in a probabilistic graphical model using a procedure called belief propagation. Here we only show example of a tree where inference is simple.



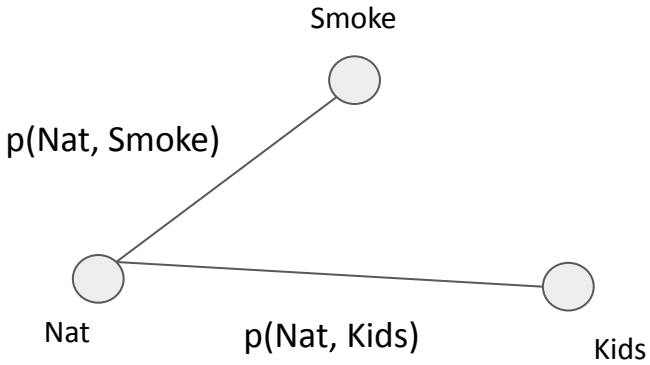
$$p(Nat = N, Smoke = S, Kids = K)$$
$$= p(Nat = N)\, p(Smoke = S \mid Nat = N)\, p(Kids = K \mid Nat = N)$$

If we measured 1-way marginal p(Nat) and 2-way marginals p(Nat, Smoke) and p(Nat, Kids) we can compute this and use it for sampling.

In the general case, we have to perform optimization to find the best parameters of the probabilistic graphical model.

Inference in a graphical model example

We can generally perform inference in a probabilistic graphical model using a procedure called belief propagation. Here we only show example of a tree where inference is simple.



$$p(Nat = N, Smoke = S, Kids = K)$$
$$= p(Nat = N)\, p(Smoke = S \mid Nat = N)\, p(Kids = K \mid Nat = N)$$

Smoke	Nat	Kids
True	EU	1
False	Other	0
True	EU	1
True	CH	2
False	EU	0
True	EU	1

Inference in a graphical model example

Using the approach described so far we can generate some new data. But this so far is not private. We have to solve the following problems to make the synthetic data generation differentially private:

Select which marginals to estimate in differentially private manner

Measure marginals with appropriate level of noise

Privacy mechanisms

Definition 3 (Gaussian Mechanism). *Let $f : \mathcal{D} \to \mathbb{R}^p$ be a vector-valued function of the input data. The Gaussian Mechanism adds i.i.d. Gaussian noise with scale σ to $f(D)$:*

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

Exponential Mechanism is useful when we have a dataset and a set of candidates, where each candidate has a score that depends on the dataset. The mechanism outputs each candidate with a probability that exponentially depends on its score:

Definition 4 (Exponential Mechanism). *Let $q : \mathcal{D} \times \mathcal{R} \to \mathbb{R}$ be quality score function and ϵ be a parameter. Then the exponential mechanism outputs a candidate $r \in \mathcal{R}$ according to the following distribution:*

$$\Pr[\mathcal{M}(D) = r] \propto \exp\left(\epsilon \cdot q(D, r)\right)$$

Selection phase

We now have to estimate MST in a differentially private manner.

We use **exponential mechanism** to estimate the maximum weight edge between two different components.

Algorithm 6: Differentially private measurement selection

Input: D (sensitive dataset), log (measurements of 1-way marginals), ρ (privacy parameter), $\mathcal{C}$ (initial set of (i,j) pairs to measure; empty by default)

Output: $\mathcal{C}$ (final set of (i,j) pairs to measure)

(1) Use `Private-PGM` to estimate all 2-way marginals $\bar{M}_{ij}$ from log

(2) Compute L_1 error between estimated 2-way marginal and actual 2-way marginal for all i,j: $q_{ij}(D) = \left\| M_{ij}(D) - \bar{M}_{ij} \right\|_1$ (this is a sensitivity 1 quantity)

(3) Let $G = (\mathcal{A}, \mathcal{C})$ be the graph where attributes are vertices and edges are pairs of attributes

(4) Let r be the number of connected components in G [6]

(5) Let $\epsilon = \sqrt{\frac{8\rho}{r-1}}$

(6) Repeat $r-1$ times

(7) Let S be the set of all attribute pairs (i,j), where i and j are in different connected components of G

(8) Select attribute pair (i,j) by running the `exponential mechanism` with quality score function q_{ij} on set S and privacy parameter ϵ.

(9) Add attribute pair (i,j) to $\mathcal{C}$

Measurement phase

We measure marginals with the appropriate level of noise

Algorithm 1: Measure Marginals

Input: D (sensitive dataset), $\mathcal{C}$ (a collection of attribute subsets), w_C (weights for each $C \in \mathcal{C}$), σ (noise scale)

Output: log (a list of noisy measurements together with metadata)

(1) Normalize weights, $w_C \leftarrow w_C / \sqrt{\sum_C w_C^2}$.

(2) For $C \in \mathcal{C}$:

(3) Calculate noisy marginal, $\tilde{\mu} = w_C M_C(D) + \mathcal{N}(0, \sigma^2 I)$

(4) Append 4-tuple $(w_C I,\ \tilde{\mu},\ \sigma,\ C)$ to measurement log

Brief look at theory tools: Renyi Differential Privacy

Renyi DP is generalization of more standard (ε, δ)-DP using the notion of Renyi divergence:

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha-1}\log \mathrm{E}_{x\sim Q}\left(\frac{P(x)}{Q(x)}\right)^\alpha$$

Definition 5 (Rényi Differential Privacy [45]). *A randomized mechanism $\mathcal{M}: \mathcal{D} \to \mathcal{R}$ satisfies (α, γ)-Rényi differential privacy (RDP) for $\alpha \geq 1$ and $\gamma \geq 0$, if for any neighboring datasets $D \sim D' \in \mathcal{D}$, we have:*

$$D_\alpha(\mathcal{M}(D) \,||\, \mathcal{M}(D')) \leq \gamma,$$

where $D_\alpha(\cdot \,||\, \cdot)$ is the Rényi divergence of order α between two probability distributions.

Renyi-DP to (ε, δ)-DP

We can compose different Renyi-DP mechanisms, and also convert Renyi-DP guarantee to (ε, δ)-DP:

Proposition 3 (Adaptive Composition of RDP Mechanisms [45]). *Let* $\mathcal{M}_1 : \mathcal{D} \to \mathcal{R}_1$ *be* (α, γ_1)*-RDP and* $\mathcal{M}_2 : \mathcal{D} \times \mathcal{R}_1 \to \mathcal{R}_2$ *be* (α, γ_2)*-RDP. Then the mechanism* $\mathcal{M} = \mathcal{M}_2(D, \mathcal{M}_1(D))$ *is* $(\alpha, \gamma_1 + \gamma_2)$*-RDP.*

Proposition 4 (RDP to DP [45]). *If a mechanism* $\mathcal{M}$ *satisfies* (α, γ)*-Rényi differential privacy, it also satisfies* $\left(\gamma + \frac{\log(1/\delta)}{\alpha - 1}, \delta\right)$*-differential privacy for all* $\delta \in (0, 1]$.

Privacy mechanisms

The following propositions state that both Gaussian and Exponential Mechanism satisfy Renyi-DP

Proposition 1 (Rényi-DP of the Gaussian Mechanism [22, 45]). *The Gaussian Mechanism applied to the function* $f : \mathcal{D} \to \mathbb{R}^p$ *satisfies* $\left(\alpha, \alpha\frac{\Delta_f^2}{2\sigma^2}\right)$*-RDP for all* $\alpha \geq 1$.

Proposition 2 (Rényi-DP of the Exponential Mechanism [12, 44]). *The Exponential Mechanism applied to the quality score function* $q : \mathcal{D} \times \mathcal{R} \to \mathbb{R}$ *satisfies* $(2\epsilon\Delta, 0)$*-DP and* $(\alpha, \alpha\frac{(2\epsilon\Delta)^2}{8})$*-RDP for all* $\alpha \geq 1$, *where* $\Delta = \max_{r \in \mathcal{R}} \Delta_{q(\cdot,r)}$ *is the maximum sensitivity of* q.

Properties of Renyi-DP

Proposition 3 (Adaptive Composition of RDP Mechanisms [45]). *Let* $\mathcal{M}_1 : \mathcal{D} \to \mathcal{R}_1$ *be* (α, γ_1)*-RDP and* $\mathcal{M}_2 : \mathcal{D} \times \mathcal{R}_1 \to \mathcal{R}_2$ *be* (α, γ_2)*-RDP. Then the mechanism* $\mathcal{M} = \mathcal{M}_2(D, \mathcal{M}_1(D))$ *is* $(\alpha, \gamma_1 + \gamma_2)$*-RDP.*

Proposition 4 (RDP to DP [45]). *If a mechanism* $\mathcal{M}$ *satisfies* (α, γ)*-Rényi differential privacy, it also satisfies* $\left(\gamma + \frac{\log(1/\delta)}{\alpha - 1}, \delta\right)$*-differential privacy for all* $\delta \in (0, 1]$.

How to evaluate synthetic data?

Given original dataset and several different synthetic datasets, how do we judge which one is better? Tao et al. (2021) propose several metrics:

- Fraction of pairs of attributes where original and synthetic data assign the same correlation level
- Classification accuracy of the downstream model trained on synthetic data (requires labels)
- Total variation distance between 1-way or 2-way marginals of the original and synthetic data

[1] Tao, Yuchao, et al. "Benchmarking differentially private synthetic data generation algorithms." *arXiv preprint arXiv:2112.09238* (2021).