Do machine learning systems meet the requirements of legal privacy standards?

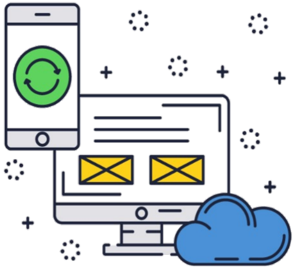
Kobbi Nissim
Georgetown University

The Data Co-Ops project https://datacoopslab.org

[PPAI, January 2023]

Machine learning systems

- The use of ML system to process personal information is growing at a rapid pace



- Credit score
- Assisting legal decision making
- Assisting hiring decisions
- Academic performance evaluation
- Online advertising and personalized content delivery

- These systems bring many benefits …
- … and also raise concerns about informational harms: privacy, discrimination and bias, misinformation, political polarization, social fragmentation …

A small sample of privacy risks in machine learning systems

- Recommendation systems [Calandrino et al. 2017]
  - Can infer information about the individual's behavior by mimicking the behavior of a target individual and then monitoring changes in a recommendation system's outputs
- Machine learning models unintentionally memorize parts of their training data and, in turn, leak secret personal information when queried [Carlini et al. 2019]
  - Auto-completion of the sentence "my social-security number is" can reveal someone's SSN
- Membership attacks [Homer et al. 2008, …]
  - given a data record and black-box access to a model, determine if the record was in the model's training dataset [Shokri et al. 2017]

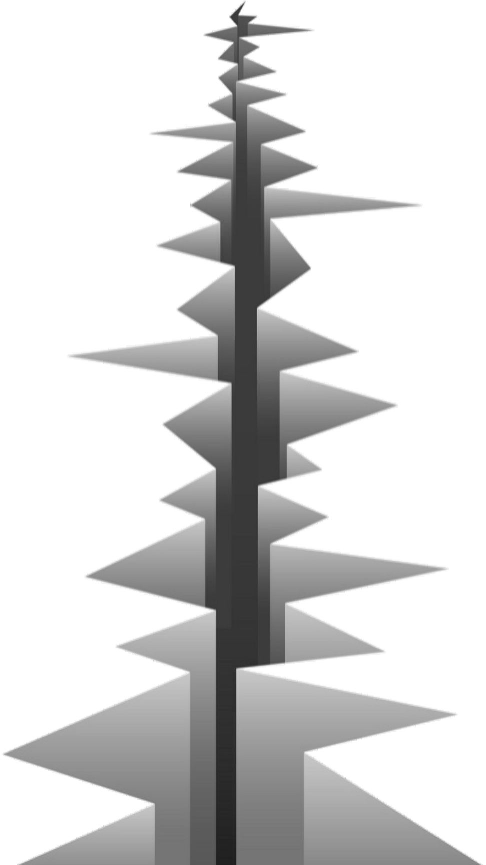
But we have …

New privacy laws

- General Data Protection Regulation
- California Consumer Privacy Act
- California Privacy Rights Act
- …

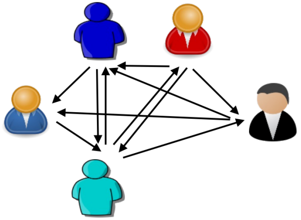
GDPR
General Data Protection Regulation

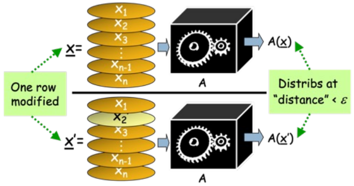
CCPA



Strong PETs

- Encryption
- Secure multiparty computing
- Differential privacy
- Blockchain
- …





Do machine learning systems meet the requirements of legal privacy standards?

Do we even understand the question?

Do machine learning systems meet legal privacy standards?

- Hard to reason about!
- Legal and technical definitions of privacy protection have evolved in diverging ways [N, Wood 2018]

- Key gaps:
  - Mathematical rigor vs. flexibility
  - Generality of protection afforded
  - Reactive vs. proactive
  - Privacy expectations vis-a-vis scientific understandings of privacy and reality of how data is used
  - Relationships to normative expectations of privacy

Do machine learning systems meet legal privacy standards?

- Design choices are frequently made opaque
  - Algorithms underlying decision support used in US courts have been considered proprietary and not subject to scrutiny [Angwin, Larson, Mattu & Kirchner 2016]

- Design of sociotechnical systems is subject to minimal regulation and oversight
  - Protections in place are widely considered to be inadequate [Barocas & Selbst 2014], [Citron & Pasquale 2014]

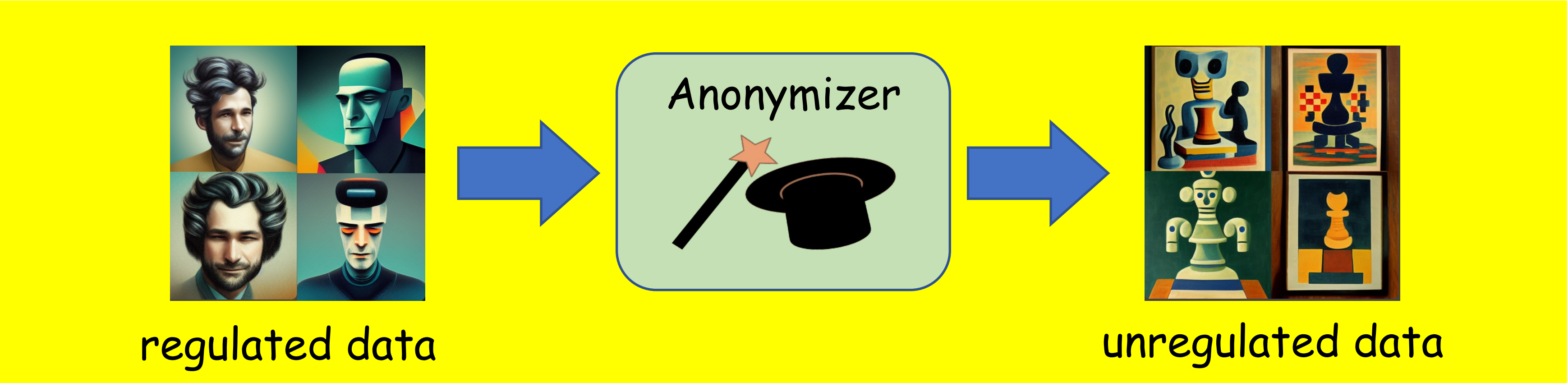
- Extremely large number of decisions are made
  - Even if only a small fraction required human review, they would quickly overwhelm judiciary or administrative systems

A concrete example:
The GDPR notion of anonymity

Based on joint work with: Micah Altman, Aloni Cohen, and Alexandra Wood

Data anonymization

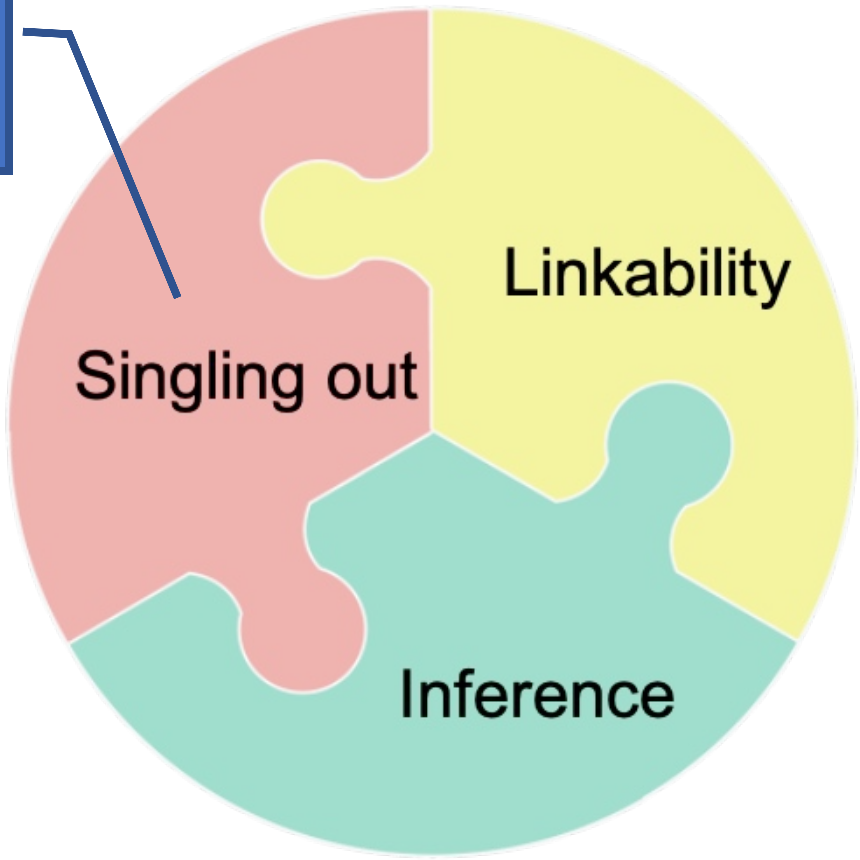
- Many privacy and data protection laws around the globe conceive of some anonymization process



- Most well-developed treatment of the concept of anonymization in regulatory guidance available today is from an opinion of the EU's Article 29 Data Protection Working Party [2014]
- The Working Party breaks down anonymization into protection from three types of attacks on unregulated (publicly released) data: linkability, singling out, and inference

Images by Midjourney

What is singling out?



Art. 29 WP general notions of attacks on released data

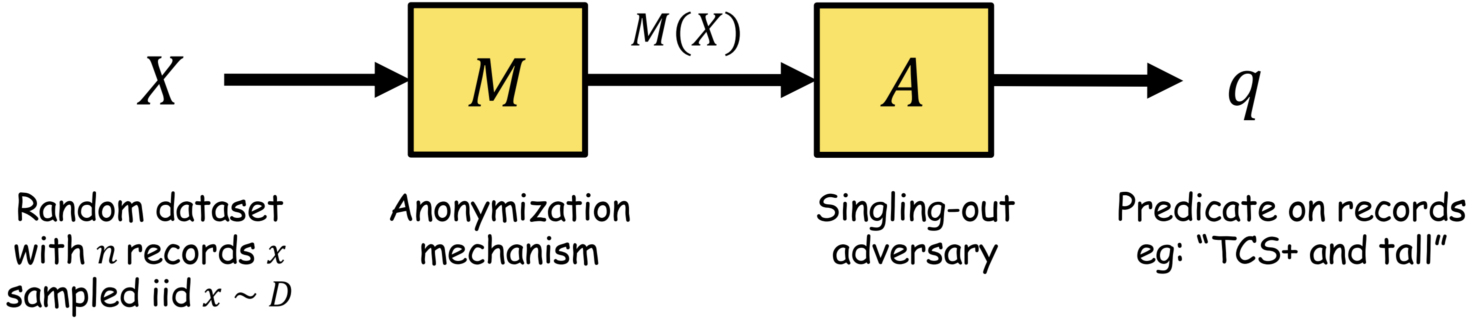
What is singling out?

- The existing A29WP guidance [2014] interprets singling out as the ability to 'isolate' an individual in the data:
  - To identify a set of attributes (or their function) that distinguishes an individual from all other individuals in the data underlying a given data release
- The guidance also lists some privacy enhancing technologies and whether they are assessed to protect against singling out

	Is Singling out still a risk?	Is Linkability still a risk?	Is Inference still a risk?
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/Tokenization	Yes	Yes	May not

Table 6. Strengths and Weaknesses of the Techniques Considered

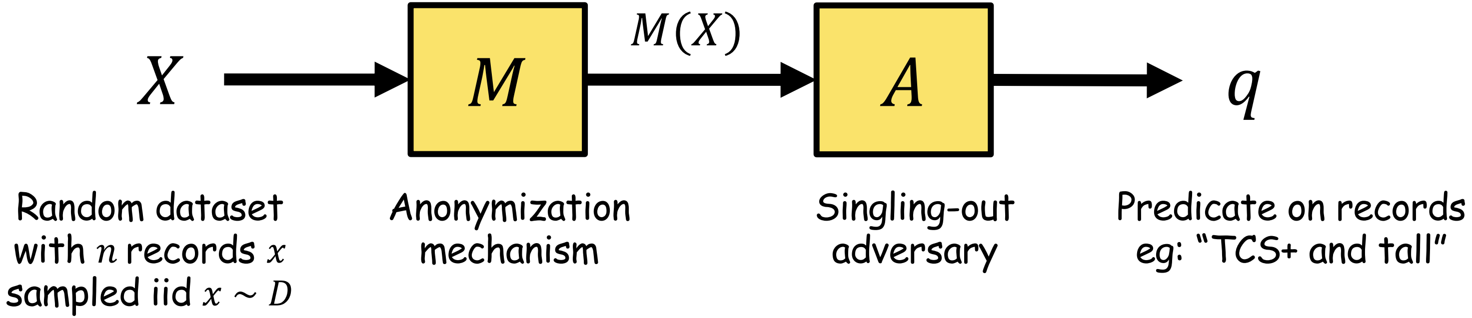
Singling out = Isolation ?



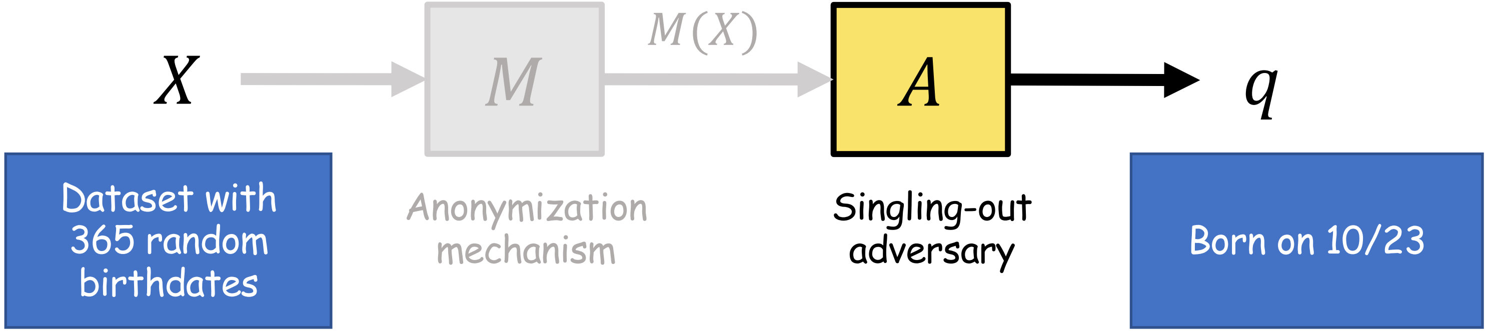
Adversary's goal: Given $M(X)$ output predicate q matching exactly 1 row in X

Definition attempt: M is secure against singling out if no adversary can isolate a row except with very small probability (over coins of X, M, A)

Every anonymization mechanism fails the isolation criterion!



Every anonymization mechanism fails the isolation criterion!

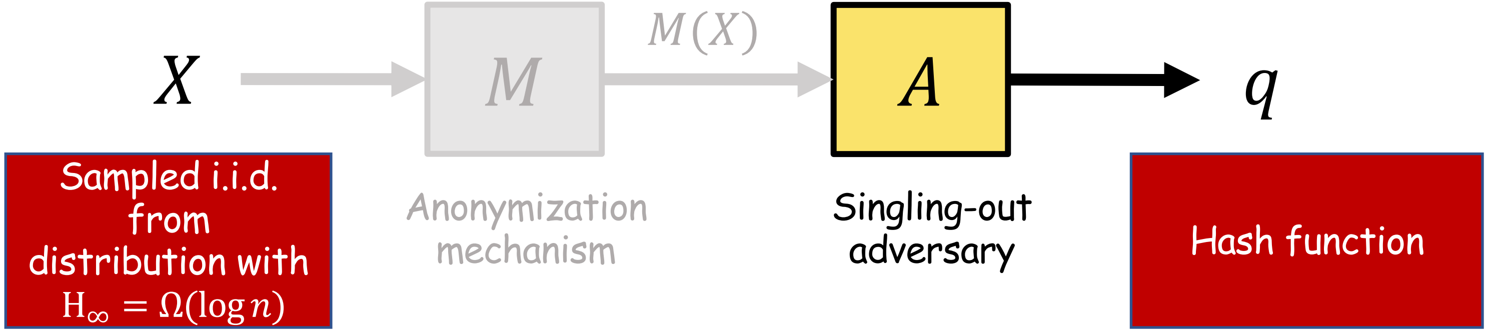


- q matches a $1/365$ fraction of the universe

$$\Pr[q^* \text{ isolates a row}] = \left(\frac{1}{365}\right)\left(1 - \frac{1}{365}\right)^{365-1} \times 365 \approx 0.37$$

- Can trivially isolate without seeing $M(X)$ and succeed with prob. $\approx\ 37\%$

Every anonymization mechanism fails the isolation criterion!

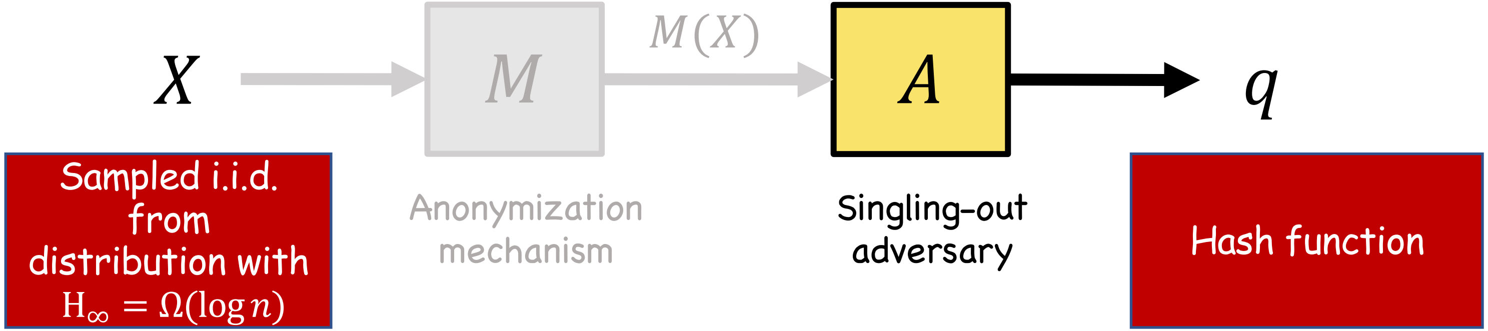


- q matches a $1/365$ fraction of the universe.

$$\Pr[q^* \text{ isolates a row}] = \left(\frac{1}{365}\right)\left(1 - \frac{1}{365}\right)^{365-1} \times 365 \approx 0.37$$

- Can trivially isolate without seeing $M(X)$ and succeed with prob. $\approx$ 37%

Every anonymization mechanism fails the isolation criterion!



- q matches a $1/n$ fraction of the universe.

$$\Pr[q^* \text{ isolates a row}] = \left(\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^{n-1} \times \approx \frac{1}{e} \approx 0.37$$

- Can trivially isolate without seeing $M(X)$ and succeed with prob. $\approx$ 37%

Can we fix the isolation criterion while preserving its spirit?

When is isolation non-trivial?

- Predicate q with $\Pr[q(x) = 1] = w$ isolates with probability $nw(1-w)^{n-1}$

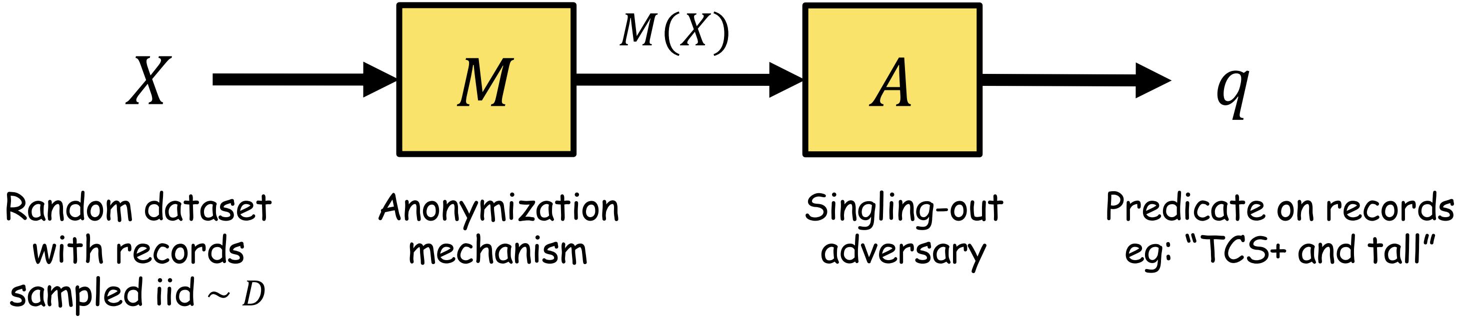
A baseline: $nw(1-w)^{n-1}$

- Idea: singling out happens when A improves significantly over the baseline

- "Born 10/23" in a dataset of 365 birthdates:
  - Attacker succeeds w.p. 37% - doable even without access to data
  - Attacker succeeds w.p. 99% - non-trivial

- "Vegan Colombian 27-year old epidemiologist, practices capoeira, loves knitting, and fluent in Dutch and Japanese"
  - Attacker succeeds with even 1% success probability – non-trivial

Security against predicate singling out [Cohen N 20]



Definition (informal): M is secure against predicate singling out attacks if there does not exist D, A s.t.

$$\Pr_{X,M,A}[A \text{ isolates with } q \text{ of weight} = negl(n)] \gg negl(n)$$

PSO security allows useful mechanisms

- Counting mechanism

$$X \Rightarrow \boxed{M_{\#g}} \Rightarrow \#x \in X \text{ satisfying } g$$

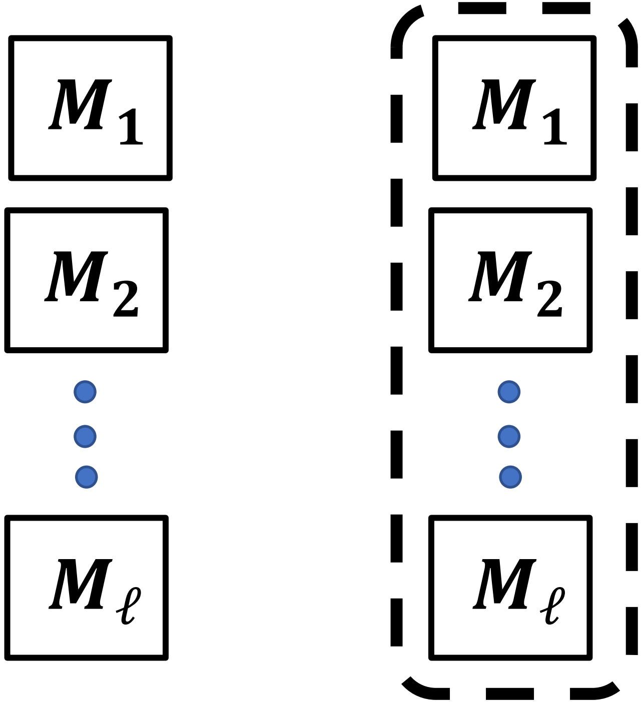
- E.g., how many people in the dataset are diabetic?

- Theorem: $M_{\#g}$ is PSO secure

Does security against PSO self-compose?

PSO secure individually

Is joint mechanism PSO secure?



Theorem [CN 20]: PSO security does not self-compose

Proof 1 utilizes $\ell = \omega(\log n)$ counting mechanisms

Proof 2 utilizes $\ell = 2$ mechanisms

Are DP and k-anonymity PSO secure?

- Theorem (informal) [CN20]: if M is d.p. then M is PSO secure
- Proof: via a connection to generalization properties of differential privacy [Dwork, Feldman, Hardt, Pitassi, Reingold, Roth '15, …]

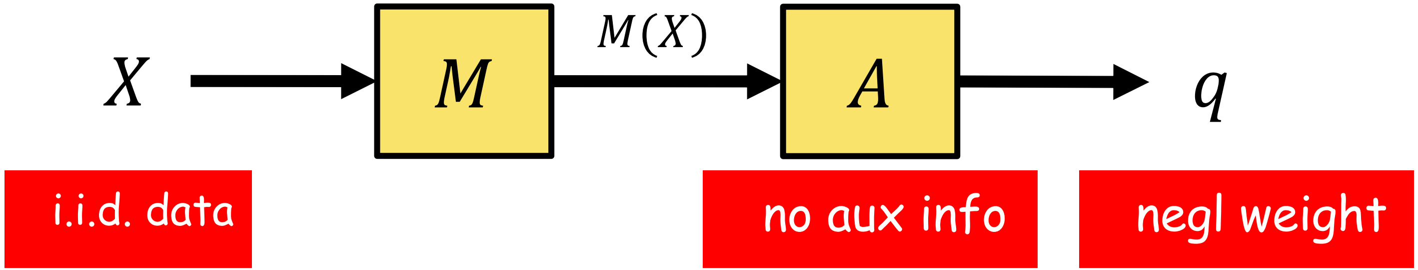
- Theorem (informal) [CN 20]: k-anonymity typically enables predicate singling out
- Proof: demonstrates that typically the k-anonymizer would do the hard work for the attacker, needs to be complemented with a trivial attacker (using leftover hash lemma)

Why should we care?

- PSO security is not the same as the GDPR notion of singling out!
- Does this mean that the use of DP satisfies the GDPR requirement wrt singling out? ?
- Does this mean that the use of k-anonymity does not satisfy the GDPR requirement wrt singling out? ?

Let's review our modeling assumptions

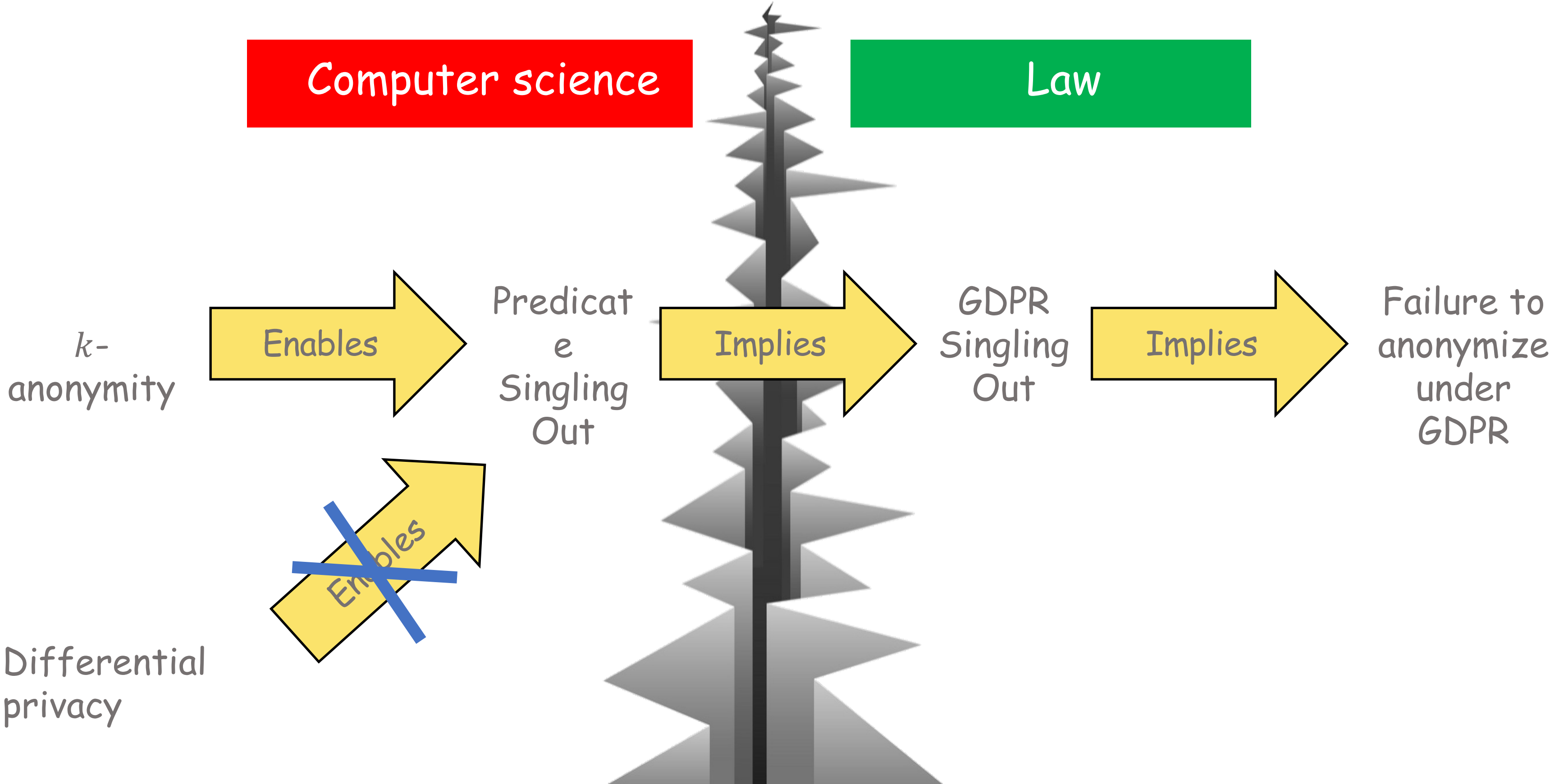
- Design choices for security against predicate singling out:



- Very likely weaker than what GDPR regulators had in mind for singling out

- Failure to protect against predicate singling out very likely implies failure to protect against GDPR singling out

A "legal theorem" for singling out



Back to the Art. 29 Working Party assessment

	Is Singling out still a risk?	Is Linkability still a risk?	Is Inference still a risk?
Pseudonymisation	Yes	Yes	Yes
Noise addition	Yes	May not	May not
Substitution	Yes	Yes	May not
Aggregation or K-anonymity	No	Yes	Yes
L-diversity	No	Yes	May not
Differential privacy	May not	May not	May not
Hashing/Tokenization	Yes	Yes	May not

Table 6. Strengths and Weaknesses of the Techniques Considered

Summary: Do machine learning systems meet the requirements of legal privacy standards?

- Difficulty in answering the question: significant gaps between regulatory and technical conceptions of privacy

- Much work needed towards bridging CS and privacy law, beyond anonymization concepts:
  - Need strategies for translating regulatory requirements into technical requirements that can be implemented in systems
  - Example privacy concepts from the regulation that need careful technical treatment: data deletion, statistical purposes, opt out, consent, …
  - Example privacy concepts from the technical literature that need to be embedded in regulation: composition, privacy budget, …