Crypto & Law (part 1)

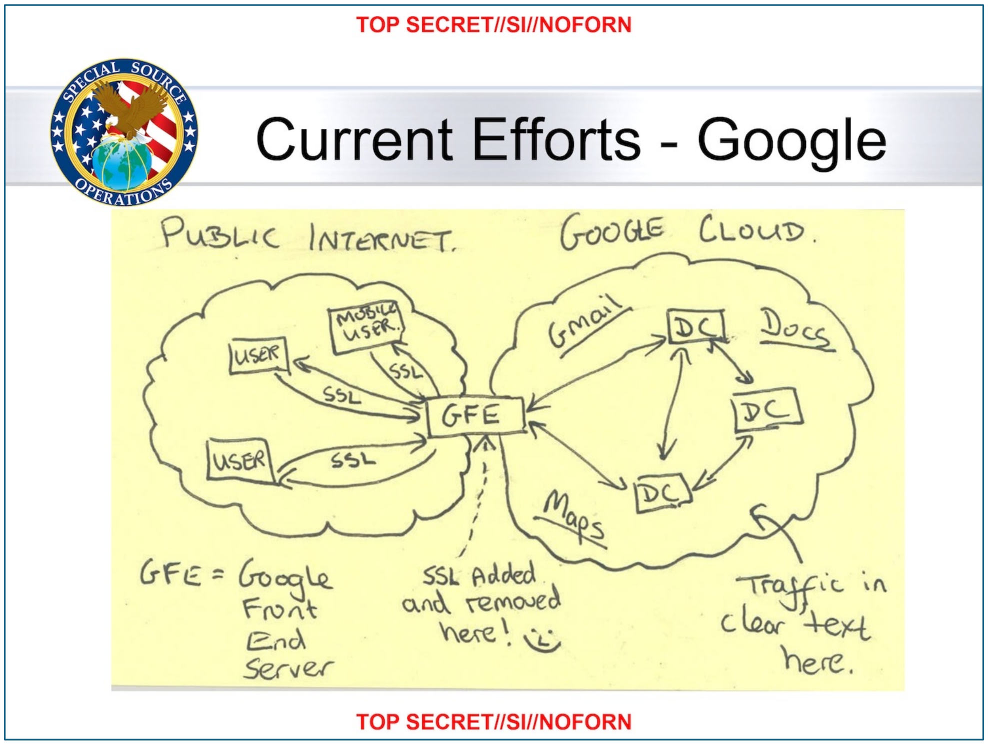
Aloni Cohen

Selected Areas in Cryptography Summer School

August, 2024

Montreal

Law of cryptography



Cryptography with legal agents / contexts

Using cryptography to understand law

Privacy Law ⟷ **Cryptography & Privacy**

Using cryptography to understand law

1. **Extract** relevant text and examples

2. **Formalize** mathematically

3. **Analyze**, alone and in relation to other notions

4. **Draw** legal conclusions

Legal analysis

Mathematical modeling & analysis

Why?

- Scale of automated decision making
- Compliance / enforcement, even in the face of change
- Learn something about the law itself
- Understand policy tradeoffs and tensions
- Exercise rights
- Steer development of new tech / law

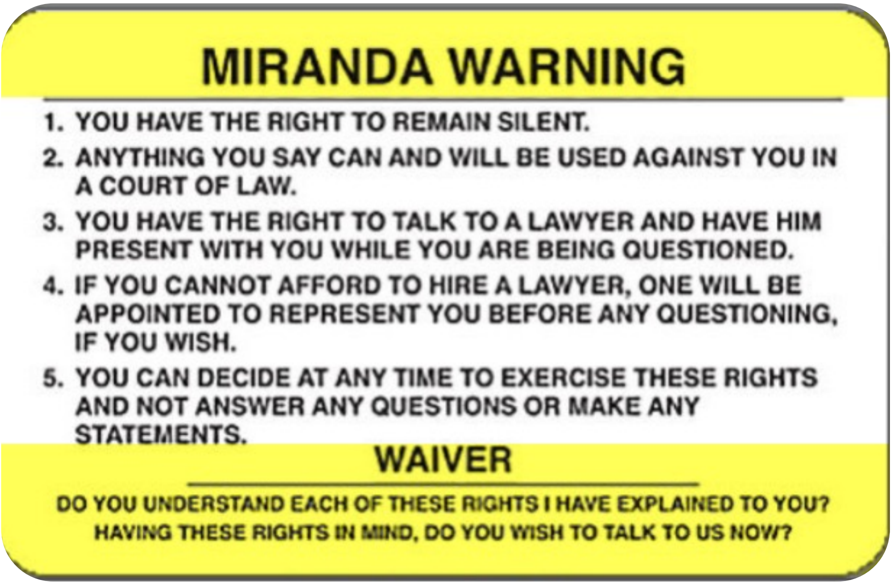
- It's fun!

Motifs

- Treating law / policy goals as first-order objectives
- Internalize law and be guided by examples
- Crypto formalisms useful, but don’t apply unthinkingly

Today


General
Data
Protection
Regulation


MIRANDA WARNING
1. YOU HAVE THE RIGHT TO REMAIN SILENT.
2. ANYTHING YOU SAY CAN AND WILL BE USED AGAINST YOU IN A COURT OF LAW.
3. YOU HAVE THE RIGHT TO TALK TO A LAWYER AND HAVE HIM PRESENT WITH YOU WHILE YOU ARE BEING QUESTIONED.
4. IF YOU CANNOT AFFORD TO HIRE A LAWYER, ONE WILL BE APPOINTED TO REPRESENT YOU BEFORE ANY QUESTIONING, IF YOU WISH.
5. YOU CAN DECIDE AT ANY TIME TO EXERCISE THESE RIGHTS AND NOT ANSWER ANY QUESTIONS OR MAKE ANY STATEMENTS.
WAIVER
DO YOU UNDERSTAND EACH OF THESE RIGHTS I HAVE EXPLAINED TO YOU?
HAVING THESE RIGHTS IN MIND, DO YOU WISH TO TALK TO US NOW?

This paper treats a class of codes made possible by restrictions on measurement related to the uncertainty principal. Two concrete examples and some general results are given.

Conjugate Coding *

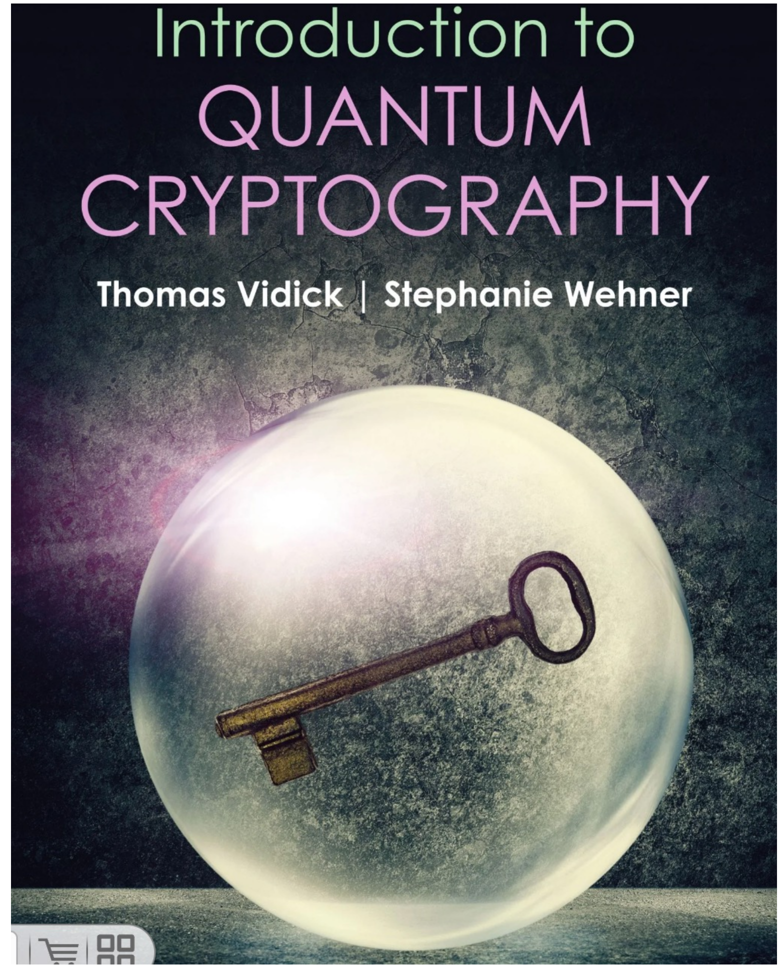
Stephen Wiesner

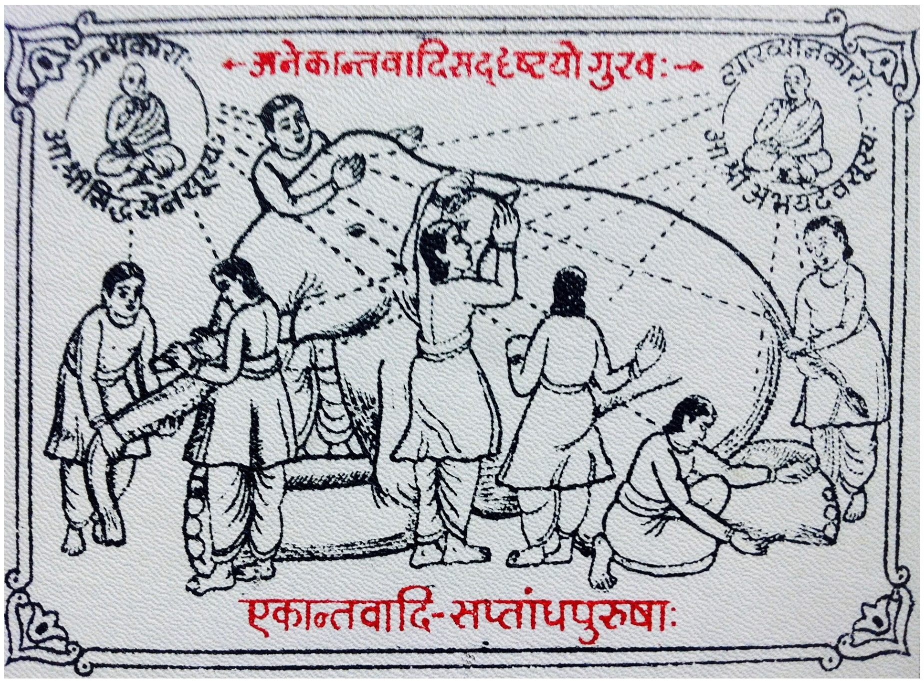
Columbia University, New York, N.Y.
Department of Physics

The uncertainty principle imposes restrictions on the capacity of certain types of communication channels. This paper will show that in compensation for this "quantum noise", quantum mechanics allows us novel forms of coding without analogue in communication channels adequately described by classical physics.

---

* Research supported in part by the National Science Foundation




अनेकान्तवादिसद्दृष्टयो गुरवः
एकान्तवादि-सप्तांधपुरुषाः


THIS IS WHERE YOU LOST YOUR WALLET?
NO, I LOST IT IN THE PARK. BUT THIS IS WHERE THE LIGHT IS.

Resources

- ACM CS&Law conference
  - https://computersciencelaw.org/
  - (First) deadline: Sept 30
  - Conference: March 2025 in Munich
- CS+Law Workshop
  - https://www.cslawworkshop.org/
  - monthly on Zoom
- GenLaw
  - https://www.genlaw.org/

How did I end up here?

I am not a lawyer…

The New York Times

Justices Say GPS Tracker Violated Privacy Rights

Share full article 290

By Adam Liptak
Jan. 23, 2012

WASHINGTON — The Supreme Court on Monday ruled unanimously that the police violated the Constitution when they placed a Global Positioning System tracking device on a suspect's car and monitored its movements for 28 days.

A set of overlapping opinions in the case collectively suggested that a majority of the justices are prepared to apply broad privacy principles to bring the Fourth Amendment's ban on unreasonable searches into the digital age, when law enforcement officials can gather extensive information without ever entering an individual's home or vehicle.

Syllabus

NOTE: Where it is feasible, a syllabus (headnote) will be released, as is being done in connection with this case, at the time the opinion is issued. The syllabus constitutes no part of the opinion of the Court but has been prepared by the Reporter of Decisions for the convenience of the reader. See *United States* v. *Detroit Timber & Lumber Co.*, 200 U. S. 321, 337.

SUPREME COURT OF THE UNITED STATES

Syllabus

UNITED STATES *v.* JONES

CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA CIRCUIT

No. 10–1259. Argued November 8, 2011—Decided January 23, 2012

The Government obtained a search warrant permitting it to install a Global-Positioning-System (GPS) tracking device on a vehicle registered to respondent Jones's wife. The warrant authorized installation in the District of Columbia and within 10 days, but agents installed the device on the 11th day and in Maryland. The Government then tracked the vehicle's movements for 28 days. It subsequently secured an indictment of Jones and others on drug trafficking conspiracy charges. The District Court suppressed the GPS data obtained while the vehicle was parked at Jones's residence, but held the remaining data admissible because Jones had no reasonable expectation of privacy when the vehicle was on public streets. Jones was convicted. The D. C. Circuit reversed, concluding that admission of the evidence obtained by warrantless use of the GPS device violated the Fourth Amendment.

Held: The Government's attachment of the GPS device to the vehicle, and its use of that device to monitor the vehicle's movements, constitutes a search under the Fourth Amendment. Pp. 3–12.

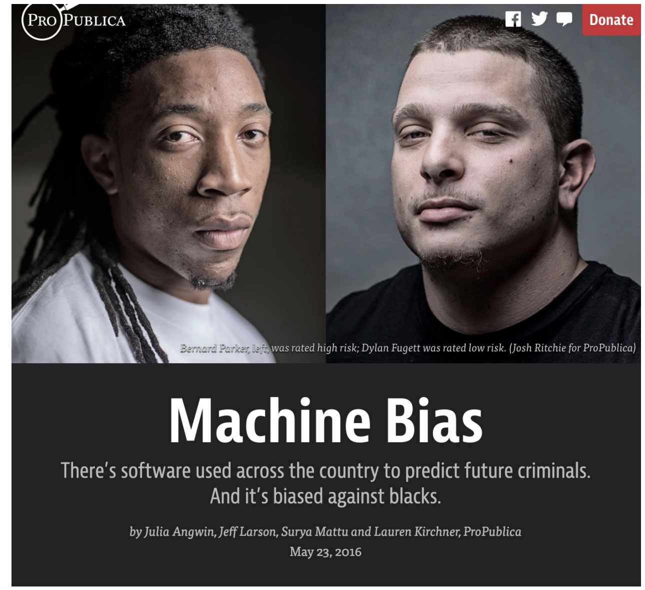
Massachusetts Institute of Technology

GEORGETOWN UNIVERSITY LAW CENTER

Georgetown University Law Center 1315/MIT 6.S978
Privacy Legislation: Law and Technology
Spring 2016

Class meetings:

@MIT:	Thursday 3:30 - 5:00 Room 9-152
@GULC:	Thursday 3:30 - 5:30 Room 200



VE RI TAS

Boston University Faculty of Computing & Data Sciences

CYBER ALLIANCE

Seminar Series

Harvard University
Privacy Tools Project

Art. 17 GDPR

Right to erasure ('right to be forgotten')

1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:

US State Privacy Legislation Tracker 2023

Comprehensive Consumer Privacy Bills

STATE	LEGISLATIVE PROCESS	STATUTE/BILL (HYPERLINKS)	COMMON NAME	CONSUMER RIGHTS									BUSINESS OBLIGATIONS				
				Right to access	Right to correct	Right to delete	Right to opt out of certain processing	Right to portability	Right to opt out of sales	Right to opt in for sensitive data processing	Right against automated decision making	Private right of action	Opt-in default (requirement age)	Notice/transparency requirement	Risk assessments	Prohibition on discrimination (exercising rights)	Purpose/processing limitation
LAWS SIGNED (TO DATE)																	
California		CCPA	California Consumer Privacy Act (2018; effective Jan. 1, 2020)	X		X		X	X			L	16	X			X
		Proposition 24	California Privacy Rights Act (2020; fully operative Jan. 1, 2023)	X	X	X	S	X	X		X	L	16	X	X	X	X
Colorado		SB 190	Colorado Privacy Act (2021; effective July 1, 2023)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Connecticut		SB 6	Connecticut Data Privacy Act (2022; effective July 1, 2023)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Indiana		SB 0005	Indiana Consumer Data Protection Act (2023; effective Jan. 1, 2026)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Iowa		SF 262	Iowa Consumer Data Protection Act (2023; effective Jan. 1, 2025)	X		X		X	X		X~		S/13	X		X	X
Montana		SB 384	Montana Consumer Data Privacy Act (2023, effective Oct. 1, 2024)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Tennessee		HB 1181	Tennessee Information Protection Act (2023; effective July 1, 2024)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Utah		SB 227	Utah Consumer Privacy Act (2022; effective Dec. 31, 2023)	X		X	P	X	X				13	X		X	
Virginia		SB 1392	Virginia Consumer Data Protection Act (2021; effective Jan. 1, 2023)	X	X	X	P	X	X	X	X~		S/13	X	X	X	X

Source:

US State Privacy Legislation Tracker 2023

Comprehensive Consumer Privacy Bills

ACTIVE BILLS															
Delaware	HB 154	Delaware Personal Data Privacy Act	X	X	X	P	X	X	X	X		S/13	X	X	X	X
Louisiana	SB 199	Louisiana Consumer Privacy Act	X	X	X	P	X	X				S/13	X	X	X	
Maine	LD 1973	Maine Consumer Privacy Act	X	X	X	IN	X	IN	X	X~		S/13	X	X	X	X
	LD 1977	Data Privacy and Protection Act	X	X	X	P	X		X		X	S/17	X	X	X	X
Massachusetts	HD 2281	Massachusetts Data Privacy Protection Act (C)	X	X	X	P	X	X			X	S/17	X	X	X	X
	SD 745		X	X	X	P	X	X			X	S/17	X	X	X	X
	HD 3263	Massachusetts Information Privacy and Security Act (C)	X	X	X	P	X	X	X	X~	L	S/13	X	X	X	X
	SD 1971		X	X	X	P	X	X	X	X~	L	S/13	X	X	X	X
	HD 3245	Internet Bill of Rights	X	X	X	P	X			X		16	X	X	X	X
New Hampshire	SB 255		X	X	X	X	X	X	X	X~		S/13	X	X	X	X
New Jersey	SB 3714	New Jersey Disclosure and Accountability Transparency Act (C)	X	X	X	X	X		X	X~	X		X	X		X
	A 505		X	X	X	X	X		X	X~	X		X	X		X
New York	A 6319	American Data Privacy and Protection Act	X	X	X	P	X	X	X		X	17	X	X	X	X
	SB 3162	(C)						X			X	13	X		X	
	A 4374							X			X	13	X		X	
	A 3593		X	X	X	IN	X			X~	X		X	X	X	X
	A 3308	Digital Fairness Act (C)	X		X	IN	X	IN		X~		ALL	X	X	X	X
	S 2277		X		X	IN	X	IN		X~		ALL	X	X	X	X
	SB 365	New York Privacy Act	X	X	X	P	X	X	X	X			X	X	X	X
	A 2587	New York Data Protection Act	X		X								X		X	X
	SB 5555	It's Your Data Act	X	X	X	IN	X	IN		X~	X	ALL	X		X	X
North Carolina	SB 525	North Carolina Consumer Privacy Act	X	X	X	P	X	X				S/13	X		X	
Oregon	SB 619		X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Pennsylvania	HB 1201	Consumer Data Privacy Act	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
	HB 708	Consumer Data Protection Act	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
Rhode Island	HB 6236	Rhode Island Data Transparency And Privacy Protection Act	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
	SB 754	Rhode Island Data Transparency and Privacy Protection Act	X	X	X	P	X	X	X	X~		S/13	X	X	X	X
	HB 5745	Rhode Island Personal Data and Online Privacy Protection Act	X	X	X	P	X	X	X	X~	X	S/13	X	X	X	X
Texas	HB 4	Texas Data Privacy and Security Act	X	X	X	P	X	X	X	X~		S/13	X	X	X	X

Source:

What does deletion from ML models require?

The ”machine unlearning” question* [CY 15, GGVZ 19, GJNRSW 21, ...]

Papers routinely conflate the question & proposed answers

"Nothing" is not the answer

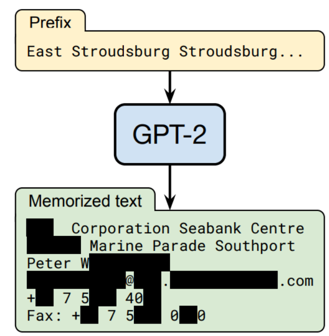
Extracting Training Data from Diffusion Models

Nicholas Carlini[*1] *Jamie Hayes*[*2] *Milad Nasr*[*1]
Matthew Jagielski[+1] *Vikash Sehwag*[+4] *Florian Tramèr*[+3]
Borja Balle[†2] *Daphne Ippolito*[†1] *Eric Wallace*[†5]

Original:
Generated:

Extracting Training Data from Large Language Models

Nicholas Carlini[1] Florian Tramèr[2] Eric Wallace[3] Matthew Jagielski[4]
Ariel Herbert-Voss[5,6] Katherine Lee[1] Adam Roberts[1] Tom Brown[5]
Dawn Song[3] Úlfar Erlingsson[7] Alina Oprea[4] Colin Raffel[1]



ML models are PII / personal data,
absent a good reason to think otherwise [VBS 18]

Making AI Forget You:
Data Deletion in Machine Learning

Antonio A. Ginart[1], Melody Y. Guan[2], Gregory Valiant[2], and James Zou[3]

Abstract

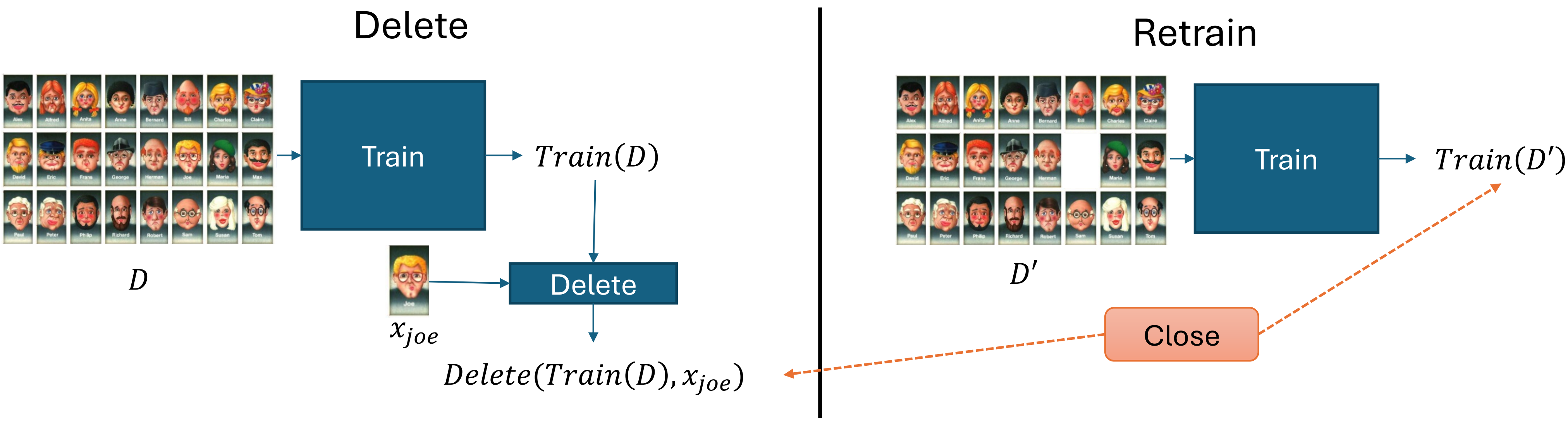
```
EMAIL -- UK BIOBANK --
Subject:  UK Biobank Application [REDACTED], Participant Withdrawal Notification [REDACTED]

Dear Researcher,

As you are aware, participants are free to withdraw form the UK Biobank at any time and request that their
data no longer be used.  Since our last review, some participants involved with Application [REDACTED]
have requested that their data should longer be used.
```

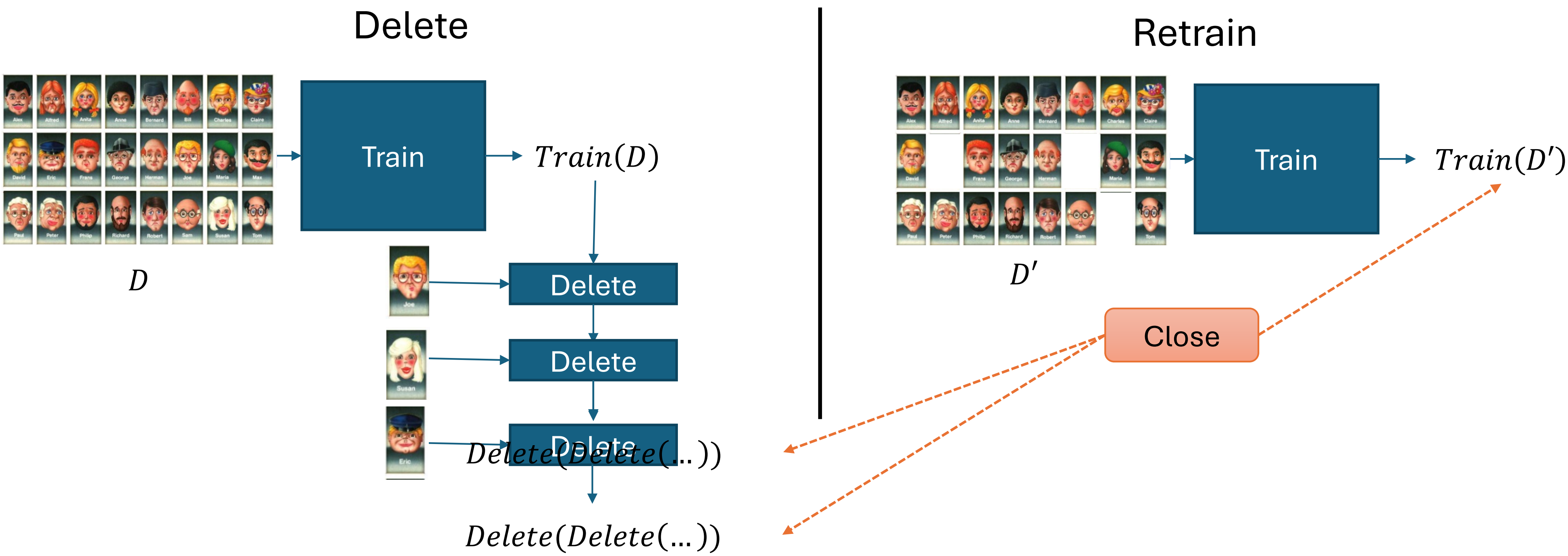
from scratch on the remaining data, which is often not computationally practical. We investigate algorithmic principles that enable efficient data deletion in ML. For the specific setting of k-means clustering, we propose two provably efficient deletion algorithms which achieve an average of over $100\times$ improvement in deletion efficiency across 6 datasets, while producing clusters of comparable statistical quality to a canonical k-means++ baseline.

History independence for unlearning



Naor, Teague. "Anti-persistence: History independent data structures" (2001)

History independence for unlearning



History independence in MUL papers: issues

- Fixable
  - Definitions often not strong enough
- More challenging
  - Tailored to ML – what about Twitter?
- The elephant in the room
  - Anonymization → users have no rights

Art. 1 GDPR

Subject-matter and objectives

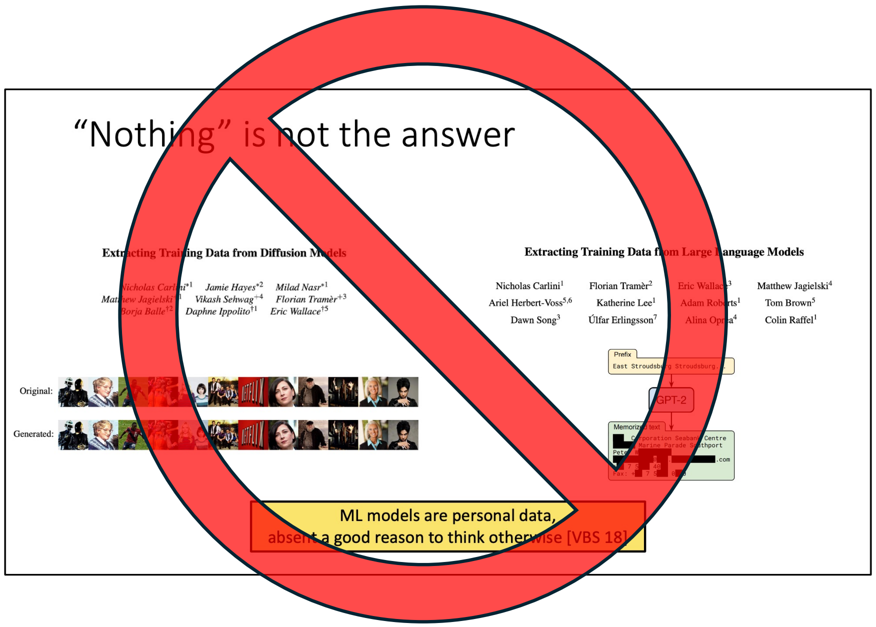
1. This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.
2. This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.

Recital 26

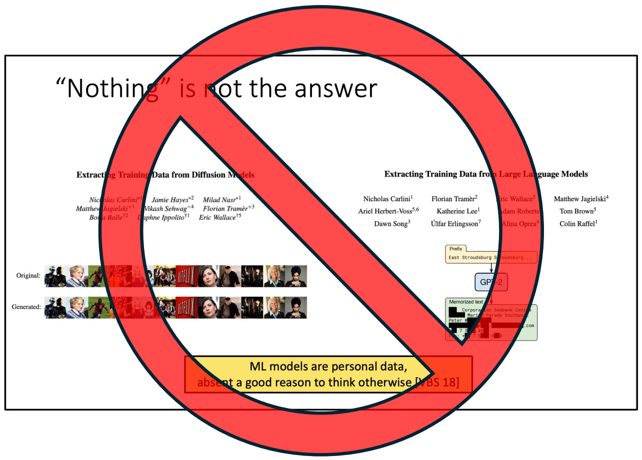
Not Applicable to Anonymous Data*

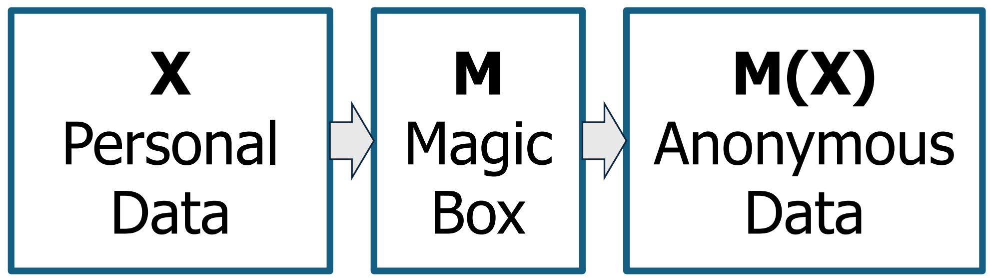
[1] The principles of data protection should apply to any information concerning an identified or identifiable natural person. [2] Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. [3] To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. [4] To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. [5] The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. [6] This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

Anonymization is all you need



Anonymization is all you need





What do we need from M for M(X) to be **anonymous** under GDPR?

In 42nd *IEEE Symposium of Security and Privacy*

Machine Unlearning

Lucas Bourtoule*‡§, Varun Chandrasekaran*†, Christopher A. Choquette-Choo*‡§, Hengrui Jia*‡§, Adelin Travers*‡§, Baiwu Zhang*‡§, David Lie‡, Nicolas Papernot‡§

University of Toronto‡, Vector Institute§, University of Wisconsin-Madison†

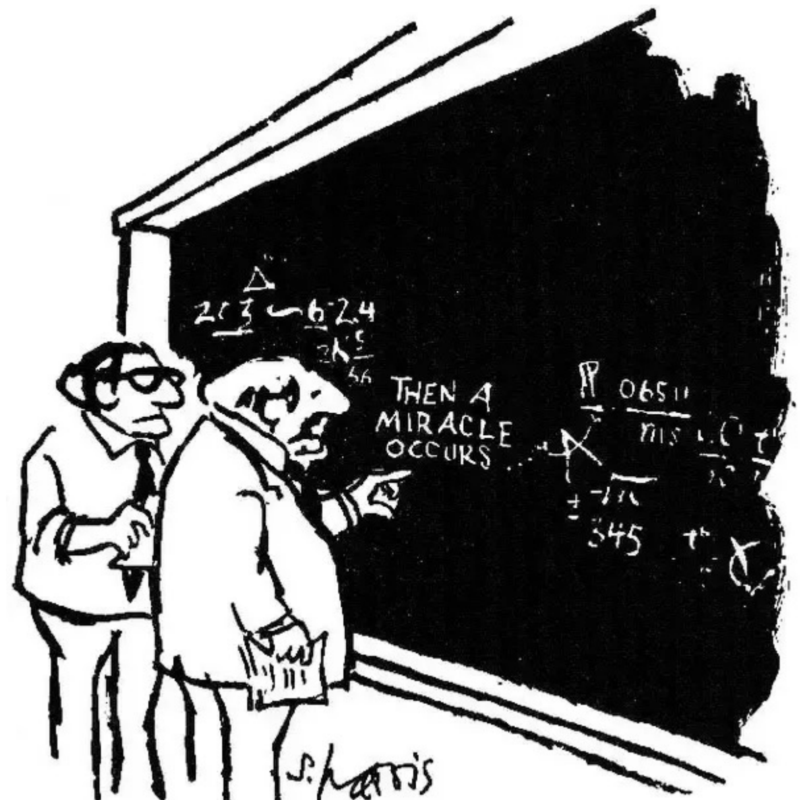
Because ML models potentially memorize training data [10], [11], it is important to unlearn what they have learned from data that is to be deleted. This problem is tangential to privacy-preserving ML—enforcing ε-differential privacy [12] with $\varepsilon \neq 0$ does not alleviate the need for an unlearning mechanism. Indeed, while algorithms which are differentially private guarantee a bound on how much individual training points contribute to the model and ensure that this contribution remains small [13], [14], there remains *a non-zero* contribution from each point. If this was not the case, the model would not be able to learn at all (see § III). In contrast, forgetting requires that *a particular* training point have *zero* contribution to the model, which is orthogonal to the guarantee provided by differential privacy.

Begs the question: does DP anonymize?


100

________ anonymizes data under GDPR

- Differential privacy
- K-anonymity / de-identification
- Synthetic data
- ML models
- Encryption
- Multiparty computation
- Federated learning
- Exact aggregates
- Noised aggregates
- Secret sharing



"I think you should be more explicit here in step two."

Hybrid concept for legal theorems

Legal Privacy Concepts

- Personally identifiable information
- De-identification
- Linkability
- Singling out
- Inference
- Data deletion

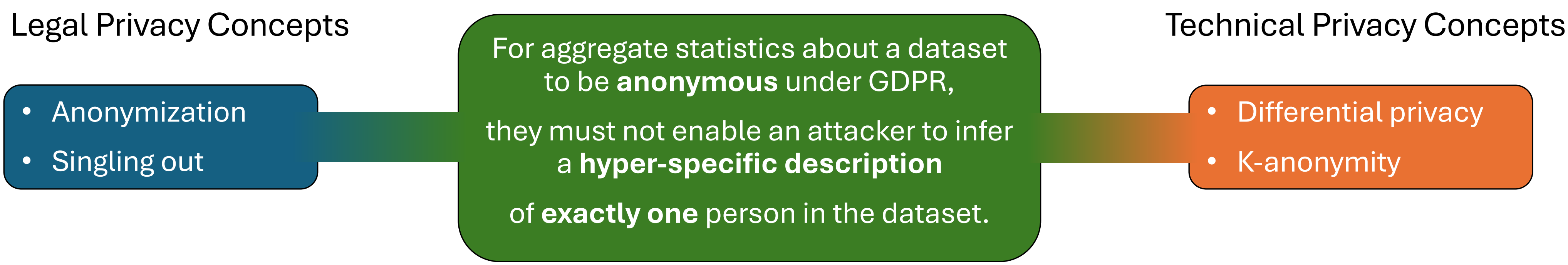
Legal interface

Tech interface

Technical Privacy Concepts

- Auxiliary information
- Post processing
- Composition
- Differential privacy
- Zero knowledge
- Secure multiparty computation
- Trust models

Predicate singling out (PSO)



Claim: Preventing PSO attacks is a **necessary** technical condition for legal anonymization under GDPR.

Theorem: Differential privacy prevents many PSO attacks.

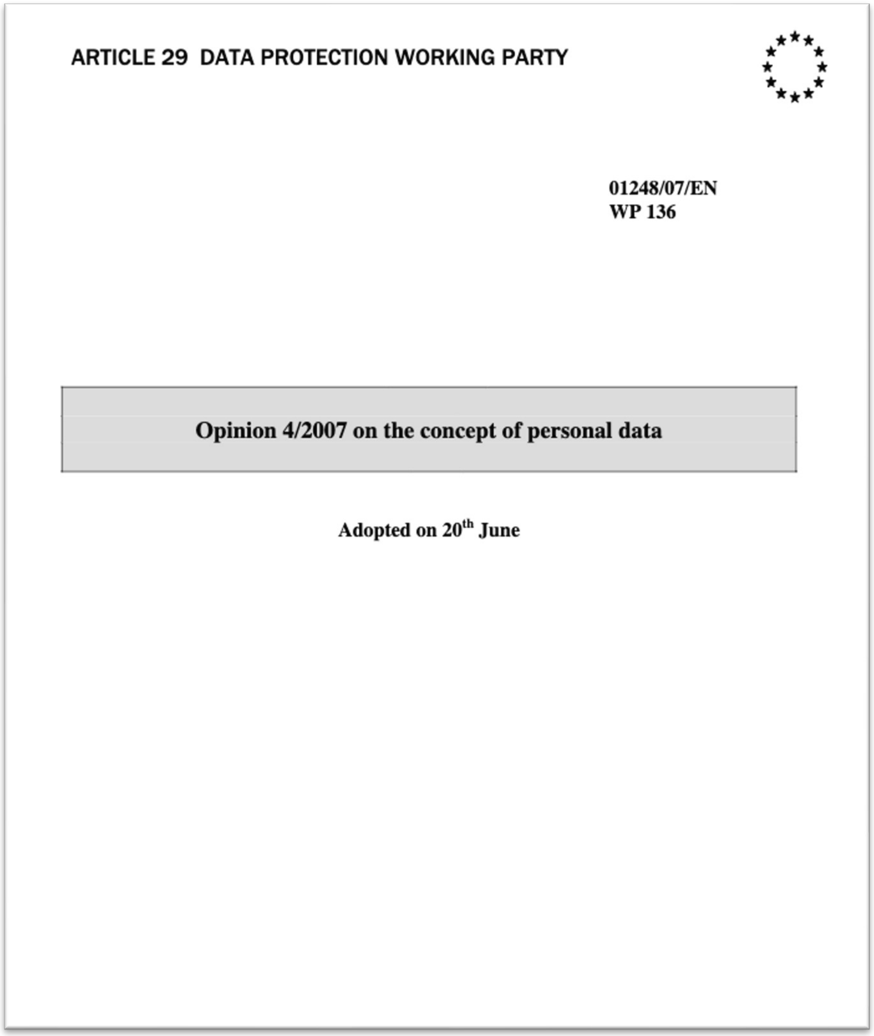
Theorem: K-anonymity enables many strong PSO attacks.

Singling out

Recital 26

Not Applicable to Anonymous Data*

[1] The principles of data protection should apply to any information concerning an identified or identifiable natural person. [2] Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. [3] To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. [4] To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. [5] The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. [6] This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.


ARTICLE 29 DATA PROTECTION WORKING PARTY

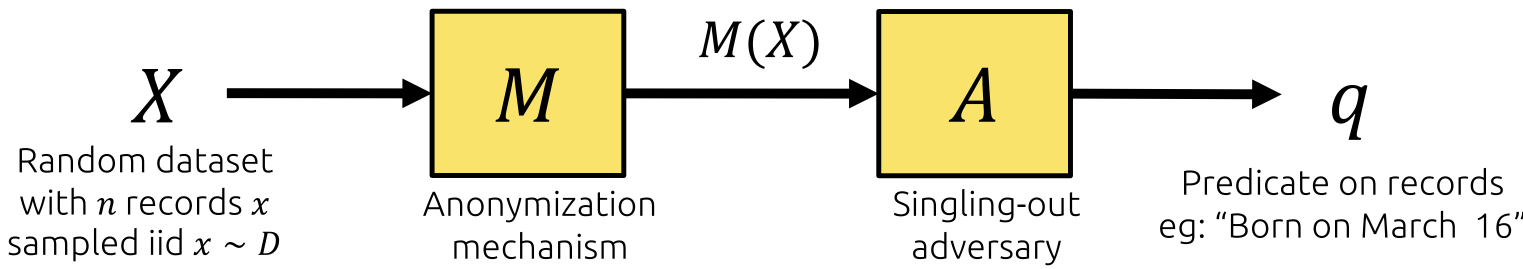
01248/07/EN
WP 136

Opinion 4/2007 on the concept of personal data

Adopted on 20th June

- A person is identified "within a group of persons [when] he or she is distinguished from all other members of the group."
- For instance, by specifying "criteria which allows him to be recognized by narrowing down the group" to a single person.

The setting



“q isolates in X” if it’s true on *exactly one* record in X

Compare A’s ability to isolate before and after seeing the output M(X)

$$X \rightarrow \boxed{M} \rightarrow \boxed{A} \rightarrow q$$

Examples, and the baseline

Isolation "q isolates in X" if it's true on *exactly one* record in X

Example ($n = 365$)

q_1 = "Born on March 16th"
q_1 isolates $\approx 37\%$ of the time

$\text{weight}(q_1) = \frac{1}{365} = \frac{1}{n}$

q_2 = "Vegan Colombian Jewish pilot fluent in Dutch"
q_2 isolates $\approx 0\%$ of the time

$\text{weight}(q_2) \approx 0$

Baseline (informal)

How often A isolates before seeing $M(X)$. Depends on weight.

Weight of q Probability of matching a random record

$$\text{weight}(q) \coloneqq \Pr_{x\sim D}[q(x)]$$

Predicate singling-out attacks (informal)

A outputs low-weight q that isolates much more often than the baseline

Calculation

$$\Pr_X[q_1 \text{ isolates in } X] = n \cdot \frac{1}{365}\left(1 - \frac{1}{365}\right)^{364} \approx e^{-1} \approx 0.37$$

$$\Pr_X[q_2 \text{ isolates in } X] \le 365 \cdot \Pr_{x\sim D}[q_2(x)] \approx 0$$

$X \to \boxed{M} \to \boxed{A} \to q$

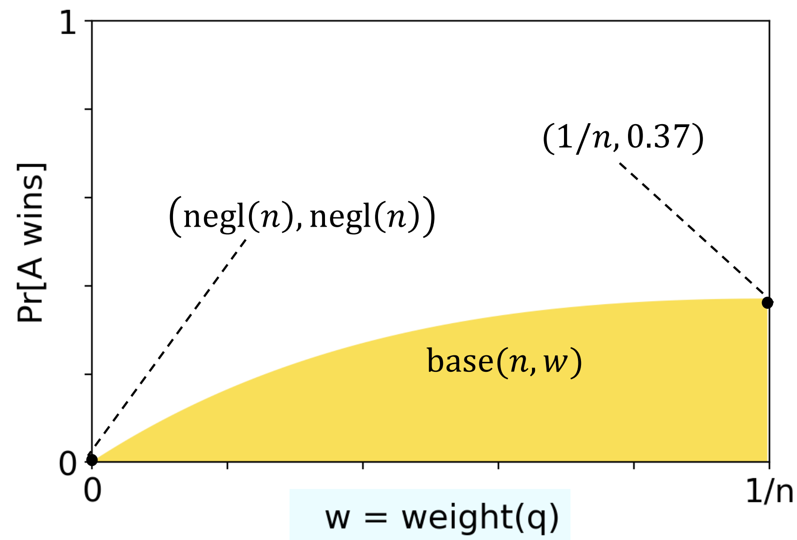
Predicate singling-out attacks [CN 20]

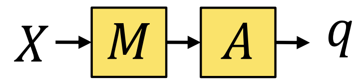
Predicate singling-out attacks (informal)

A outputs low-weight q that isolates much more often than the baseline

"A wins" for weight w $\quad (\mathrm{weight}(q) < w)$ AND (q isolates in X)

Baseline $\quad \mathrm{base}(n, w) \coloneqq \max_{A \text{ ignoring } M} \Pr_{X,M,A}[A \text{ wins}]$





Predicate singling-out attacks [CN 20]

Predicate singling-out attacks (informal)

A outputs low-weight q that isolates much more often than the baseline

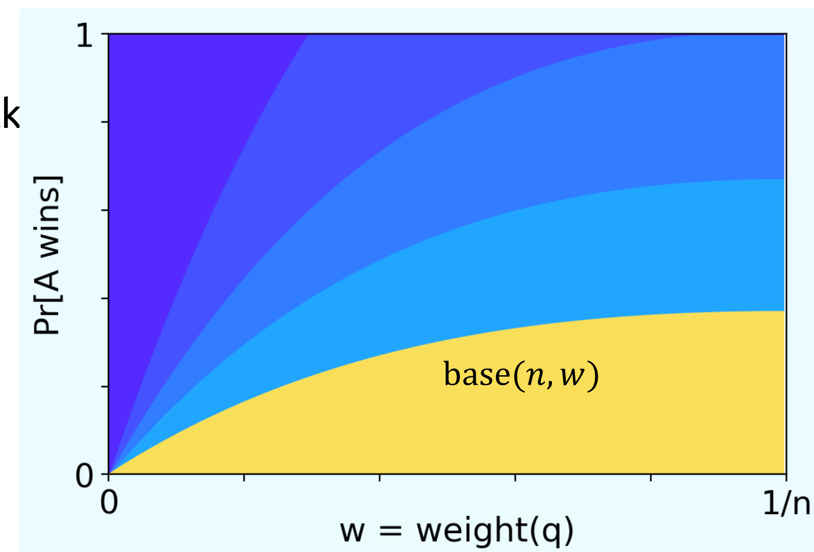
"A wins" for weight w $\quad (\mathrm{weight}(q) \; < \; w) \text{ AND } (q \text{ isolates in } X)$

Baseline $\quad \mathrm{base}(n, w) \coloneqq \max_{A \text{ ignoring } M} \Pr_{X,M,A}[A \text{ wins}]$

Definition (Predicate singling-out attack

For $w < 0 \le \frac{1}{n}$, *M enables predicates singling-out attacks* if there exist adversary A, distribution D such that

$\Pr_{X,M,A}[A \text{ wins}] \gg \mathrm{base}(n, w)$



Summary of PSO results

Theorem: For M computing exact counts

$$\Pr[A\ wins] \leq (n+1) \cdot \mathrm{base}(n, w)$$

For $\Pr[A\ wins] < 0.01$:

- Counts: $w < \frac{c}{n^2}$
- DP: $w < \frac{c}{\epsilon n}$

Theorem: For M (ϵ, δ)-DP, $w < \frac{1}{n}$

$$\Pr_{X,M,A}[A \text{ wins}] \leq (2+\epsilon) \cdot \mathrm{base}(n, w) + n\delta$$

Theorem (informal): PSO-security doesn't compose

Theorem (informal): k-anonymity enables PSO attacks

Example: Counting Mechanism

$$X \longrightarrow \boxed{M_{\#h}} \longrightarrow h(X) = \Pr_{x \leftarrow X}[h(x) = 1]$$

Theorem: For M computing exact counts

$$\Pr[A\ wins] \leq (n+1) \cdot \mathrm{base}(n, w)$$

Proof:

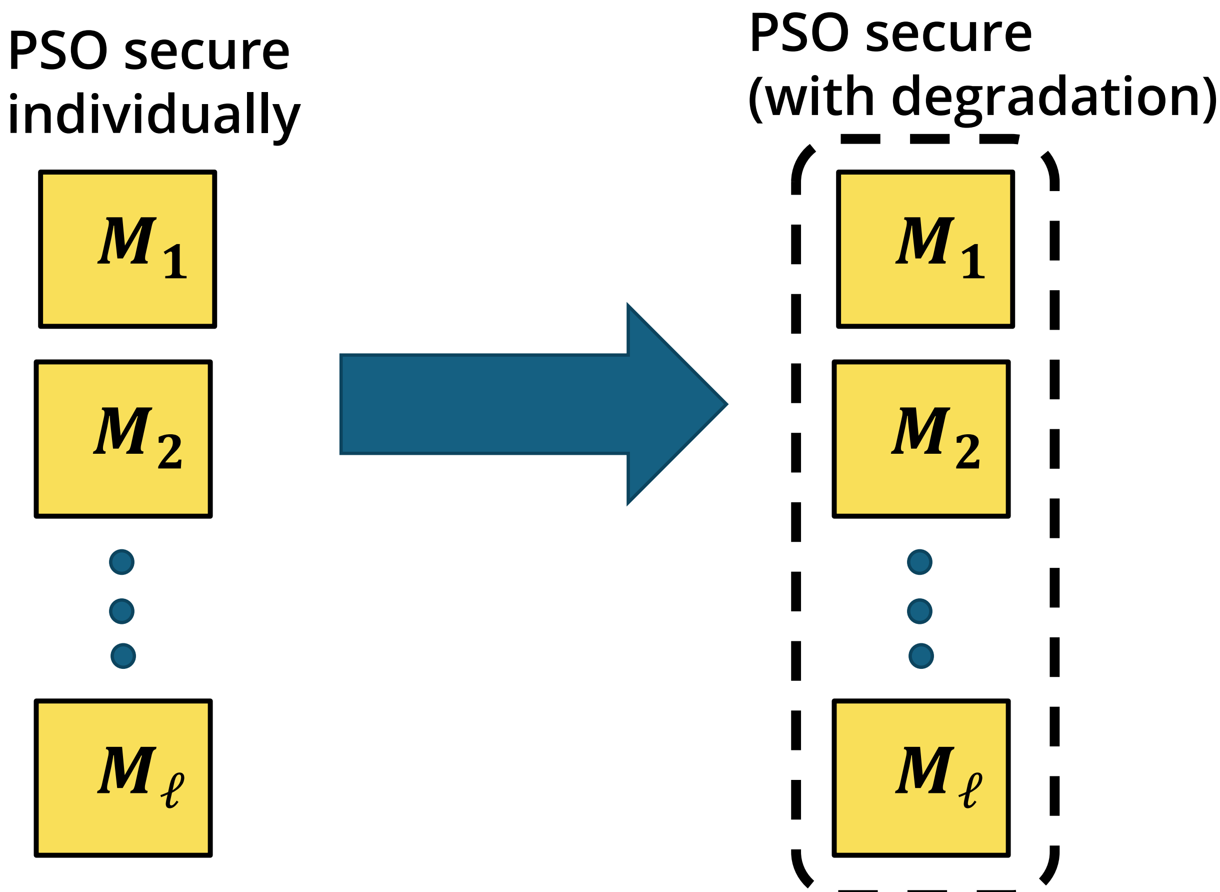
Possible answers: $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, 1\}$

Baseline attacker guesses $M_{\#h}(X)$, and runs A.

$\Rightarrow base(n, w) \geq \frac{\Pr[A\ wins]}{n+1}$

PSO security $\not\Rightarrow$ Differential privacy

Composition



Theorem

PSO security
does not compose
$\ell = 2$ times.

<u>This talk:</u> $\ell = \omega(\log n)$

Non-composition proof

Counting Mechanisms

$\boldsymbol{M_{\#h_1}}$

$\boldsymbol{M_{\#h_2}}$

$\vdots$

$\boldsymbol{M_{\#h_\ell}}$

If h_1 isolates row x,
can learn $x[1]$, $x[2]$, …

$x[1]$	$x[2]$	$x[3]$	$x[4]$	…			
0	1	0	0	1	0	0	0
0	1	1	1	0	1	1	0
1	0	0	0	0	1	1	1
1	0	1	0	1	0	1	0
0	0	0	0	1	1	1	1

X

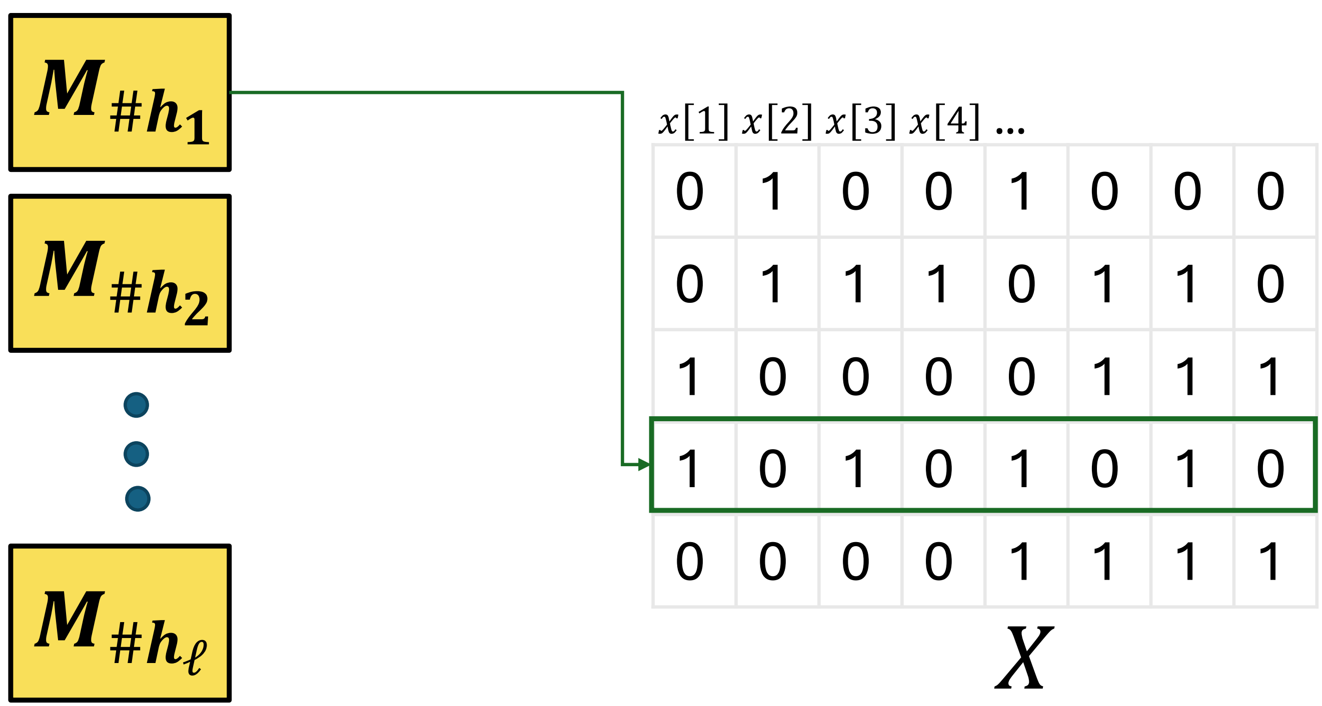
Probability ≈ 0.37!

Non-composition proof

Counting Mechanisms

If h_1 isolates row x, can learn $x[1]$, $x[2]$, ...

Probability ≈ 0.37!



Non-composition proof

Counting
Mechanisms

$\boldsymbol{M_{\#h_1}}$

$\boldsymbol{M_{\#h_2}}$ $\boldsymbol{h_1 \wedge x[1] == 1}$

$\boldsymbol{M_{\#h_\ell}}$

If h_1 isolates row x,
can learn $x[1]$, $x[2]$, ...

$x[1]$	$x[2]$	$x[3]$	$x[4]$	...			
0	1	0	0	1	0	0	0
0	1	1	1	0	1	1	0
1	0	0	0	0	1	1	1
1	0	1	0	1	0	1	0
0	0	0	0	1	1	1	1

X

Probability ≈ 0.37!

Non-composition proof

Counting Mechanisms

$\boldsymbol{M_{\#h_1}}$

$\boldsymbol{M_{\#h_2}}$

$\vdots$

$\boldsymbol{M_{\#h_\ell}}$

$\boldsymbol{h_1 \wedge x[1] == 1}$

$\boldsymbol{h_1 \wedge x[2] == 1}$

If h_1 isolates row x,
can learn $x[1]$, $x[2]$, …

$x[1]$	$x[2]$	$x[3]$	$x[4]$	…			
0	1	0	0	1	0	0	0
0	1	1	1	0	1	1	0
1	0	0	0	0	1	1	1
1	0	1	0	1	0	1	0
0	0	0	0	1	1	1	1

X

Probability ≈ 0.37!

Non-composition proof

Counting Mechanisms

$\boldsymbol{M}_{\#\boldsymbol{h}_1}$

$\boldsymbol{M}_{\#\boldsymbol{h}_2}$

$\boldsymbol{M}_{\#\boldsymbol{h}_\ell}$

$\boldsymbol{h}_1 \wedge \boldsymbol{x}[\mathbf{1}] == \mathbf{1}$

$\boldsymbol{h}_1 \wedge \boldsymbol{x}[\mathbf{2}] == \mathbf{1}$

$\boldsymbol{h}_1 \wedge \boldsymbol{x}[\mathbf{3}] == \mathbf{1}$

If h_1 isolates row x,
can learn $x[1]$, $x[2]$, …

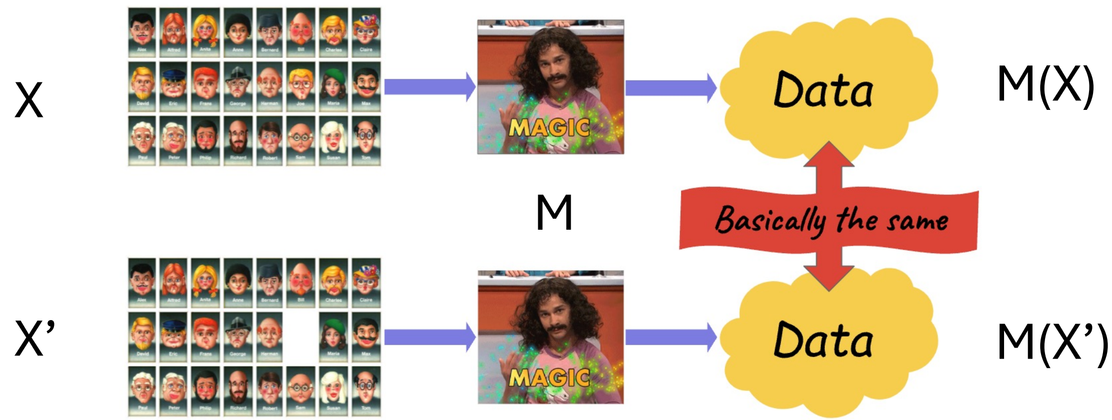
$x[1]$	$x[2]$	$x[3]$	$x[4]$	…			
0	1	0	0	1	0	0	0
0	1	1	1	0	1	1	0
1	0	0	0	0	1	1	1
1	0	1	0	1	0	1	0
0	0	0	0	1	1	1	1

X

Probability ≈ 0.37!

After ℓ bits, weight $2^{-\ell}$

Differential privacy



Definition: Random variables A and B over Ω are (ϵ, δ)-close if $\forall S \subseteq \Omega$,

$$A \approx_{\epsilon,\delta} B \qquad \Leftrightarrow \qquad \Pr[A \in \Omega] \leq e^{\epsilon} \cdot \Pr[B \in \Omega] + \delta$$

Definition: M is (ϵ, δ)-differentially private if for all X, X' differing in one item,

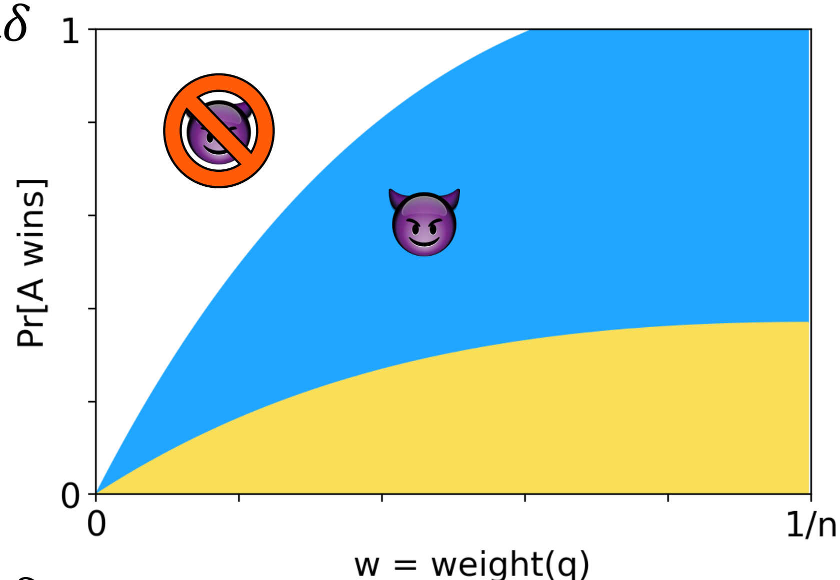
$$M(X) \approx_{\epsilon,\delta} M(X')$$

https://desfontain.es/blog/differential-privacy-awesomeness.html

Differential privacy & PSO

Theorem: For M (ϵ, δ)-DP, $w < \frac{1}{n}$

$$\Pr_{X,M,A}[A \text{ wins}] \le (2+\epsilon) \cdot \text{base}(n, w) + n\delta$$



Proof idea:

PSO attack is a type of overfitting

$$q(X) = \frac{1}{n} > w = q(D)$$

DP prevents overfitting.

$$\mathop{\mathrm{E}}_{\substack{X \sim D^n \\ h \leftarrow A \circ M(X)}}[q(X)] \le e^{\epsilon} \cdot \mathop{\mathrm{E}}_{\substack{X \sim D^n \\ h \leftarrow A \circ M(X)}}[q(D)] + \delta$$

k-anonymity

ZIP	Rich	Retired
02446	1	1
02446	0	0
02445	1	0
91011	0	0
91301	0	0
91640	1	0

M

3-Anon

ZIP	Rich	Retired
0244*	*	*
0244*	*	*
0244*	*	*
91***	*	0
91***	*	0
91***	*	0

Hierarchical Attributes generalized along a hierarchy H
(e.g., 02446 → 0244* → 024** → 02*** → 0*** → *****)

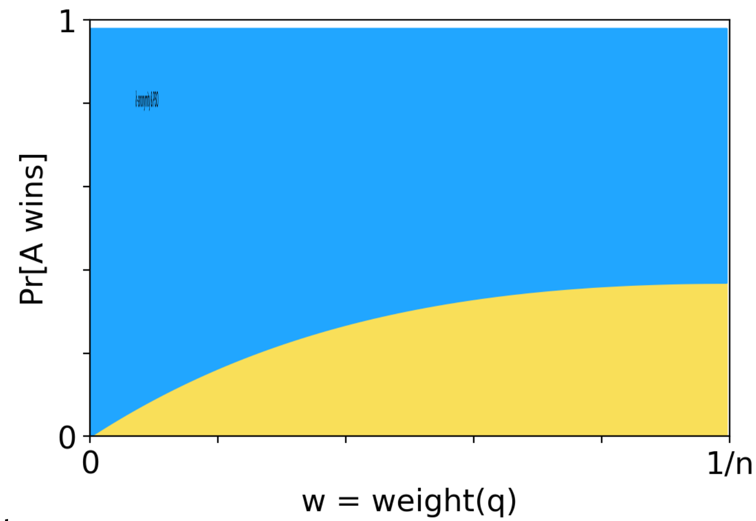
Minimal As detailed as possible along H
(e.g., Don't use 02*** when 0244* works)

k-anonymity & PSO

Theorem (Informal)

Minimal hierarchical k-anonymous mechanisms
enable **strong** predicate singling-out attacks ***against every row!***

$$X \to \boxed{M} \to \boxed{A} \to q_1, q_2, \dots, q_n$$

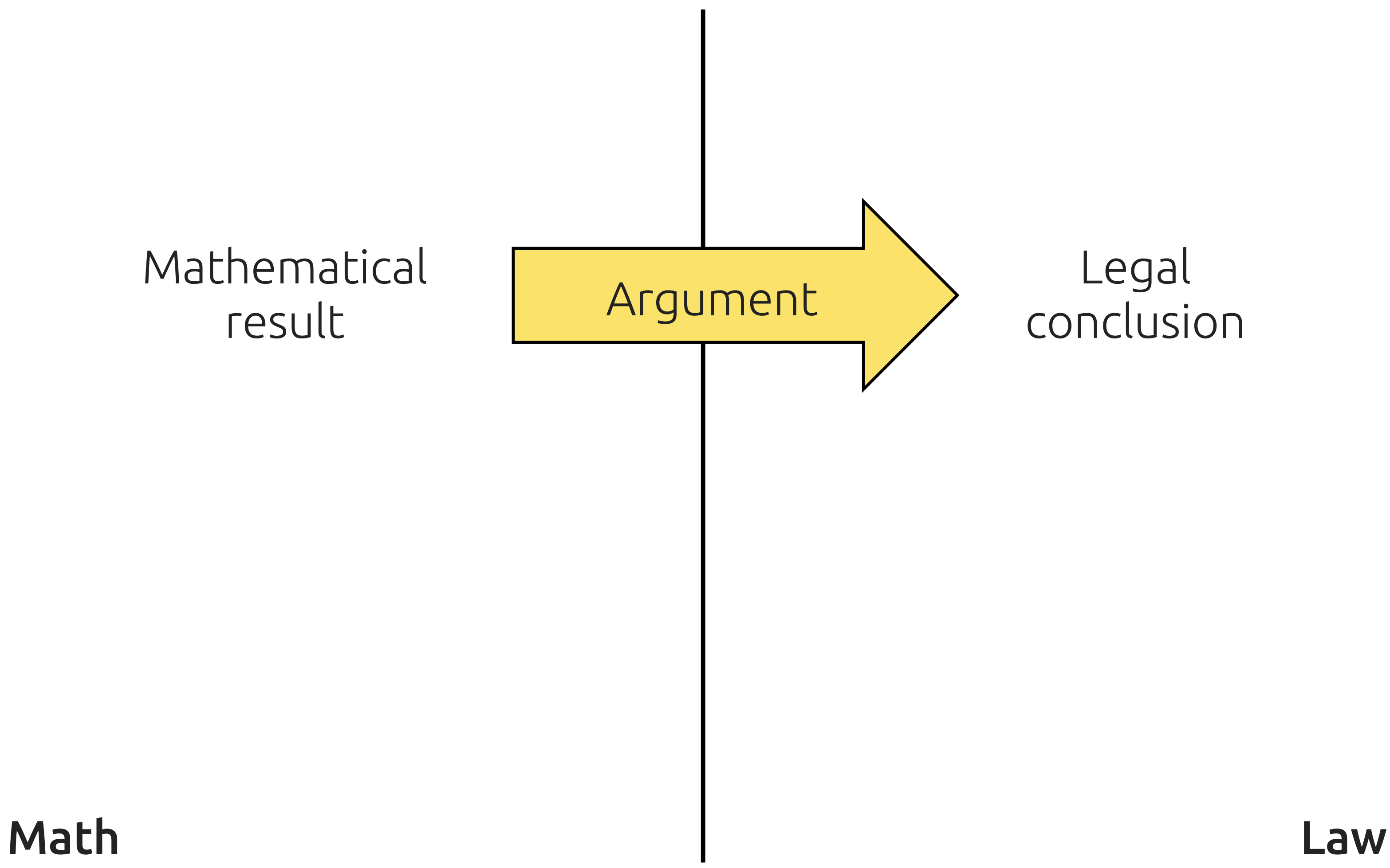


Theorem

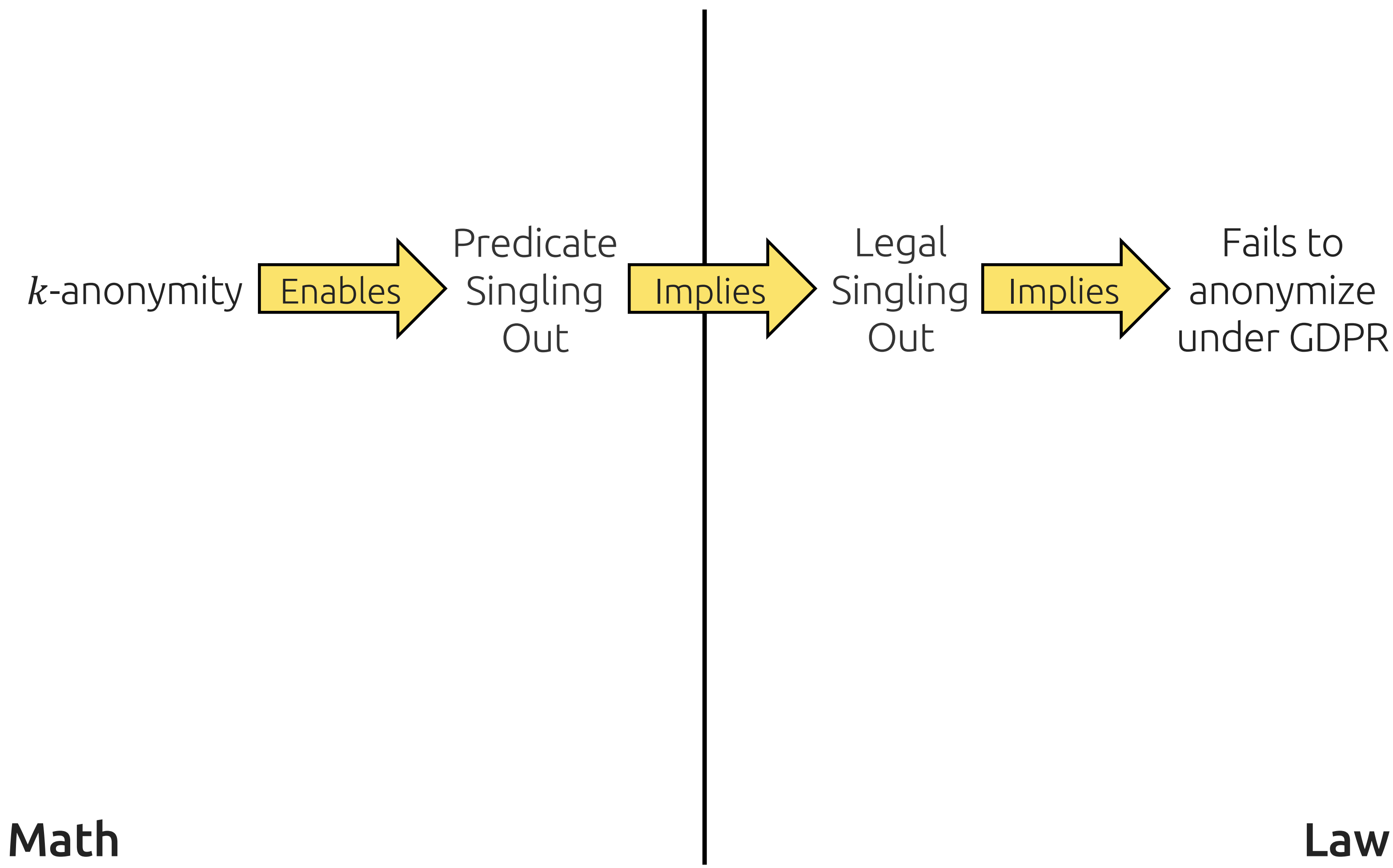
For all $k > 1$, $\alpha > 0$, weight $w < \mathrm{negl}(n)$ there exists
A, D, H such that for all minimal hierarchical k-anonymous M

$$\Pr_{X,M,A}[A \text{ wins simultaneously with every } q_i] > 1 - \alpha$$

Hybrid mathematical-legal theorem



Hybrid mathematical-legal theorem



Resolving disagreement with legal guidance

ARTICLE 29 DATA PROTECTION WORKING PARTY

	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

Opinion 05/2014 on Anonymisation Techniques

Adopted on 10 April 2014

Resolving disagreement with legal guidance

Option 1:
Legal postulate

Guidance is correct by **fiat**.

Option 2:
Squishy guidance

Guidance is **typically** correct, but allows **exceptions**.

Option 3:
Hybrid conjecture

Guidance is **best guess** at the time, can be **wrong**

updated guidance coming … eventually?

Art. 17 GDPR

Right to erasure (‘right to be forgotten’)

1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:

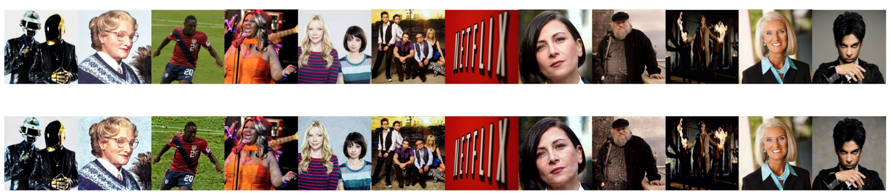
"Nothing" is not the answer

Extracting Training Data from Diffusion Models

Nicholas Carlini[*1] *Jamie Hayes*[*2] *Milad Nasr*[*1]
Matthew Jagielski[+1] *Vikash Sehwag*[+4] *Florian Tramèr*[+3]
Borja Balle[†2] *Daphne Ippolito*[†1] *Eric Wallace*[†5]

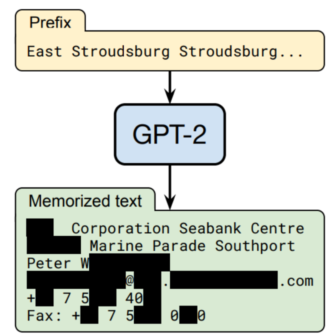
Original:

Generated:



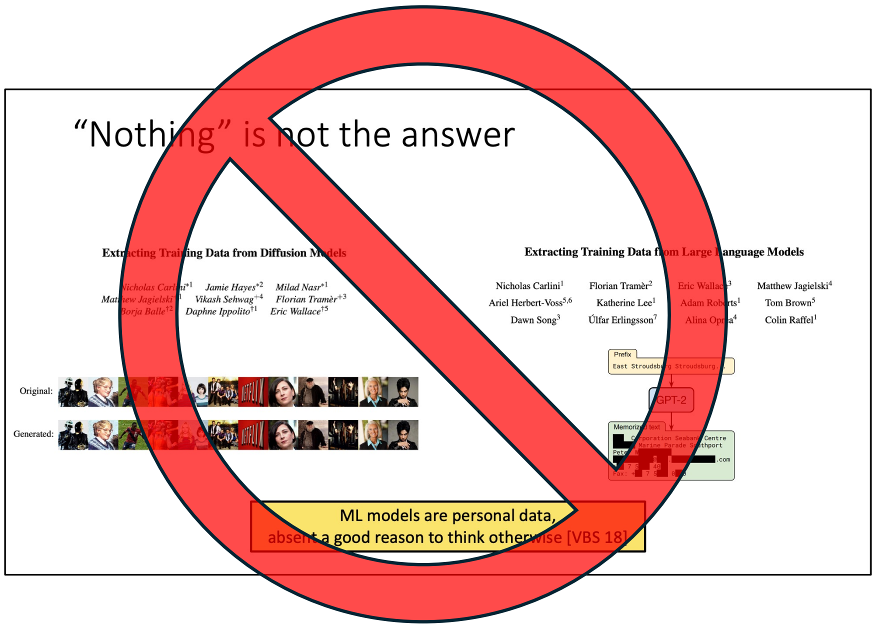
Extracting Training Data from Large Language Models

Nicholas Carlini[1] Florian Tramèr[2] Eric Wallace[3] Matthew Jagielski[4]
Ariel Herbert-Voss[5,6] Katherine Lee[1] Adam Roberts[1] Tom Brown[5]
Dawn Song[3] Úlfar Erlingsson[7] Alina Oprea[4] Colin Raffel[1]



ML models are PII / personal data,
absent a good reason to think otherwise [VBS 18]

Anonymization is all you need for erasure?

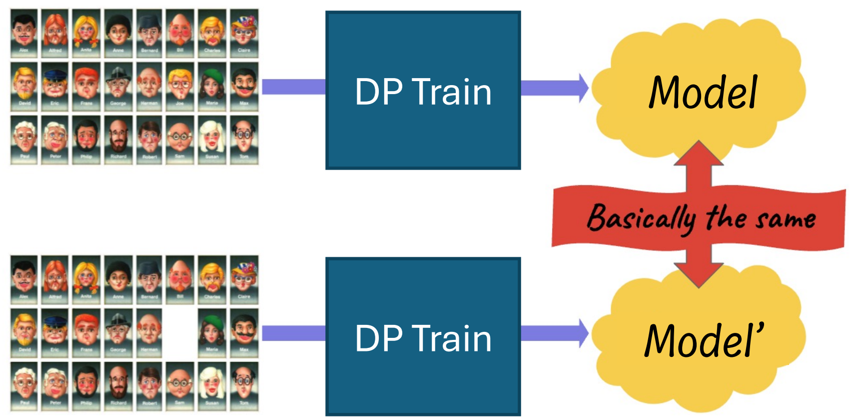


What does deletion from ML models require?

The ”machine unlearning” question* [CY 15, GGVZ 19, GJNRSW 21, ...]

**Papers routinely conflate the question & proposed answers*

Differential privacy for unlearning



Definition: Random variables A and B over Ω are (ϵ, δ)-close if $\forall S \subseteq \Omega$,

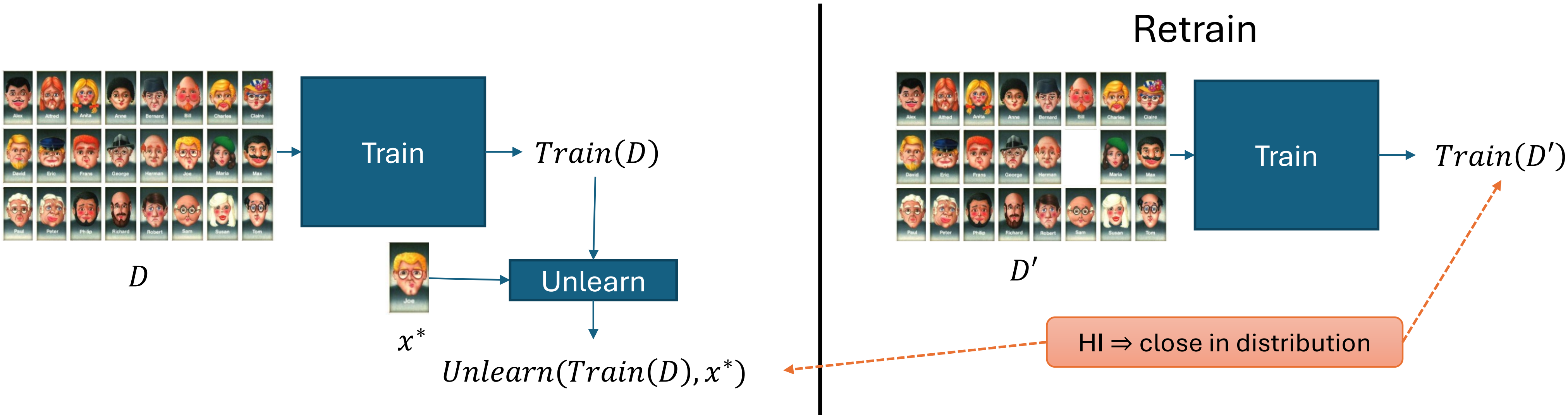
$$A \approx_{\epsilon,\delta} B \quad \Leftrightarrow \quad \Pr[A \in \Omega] \leq e^{\epsilon} \cdot \Pr[B \in \Omega] + \delta$$

Definition: M is (ϵ, δ)-differentially private if for all X, X' differing in one item,

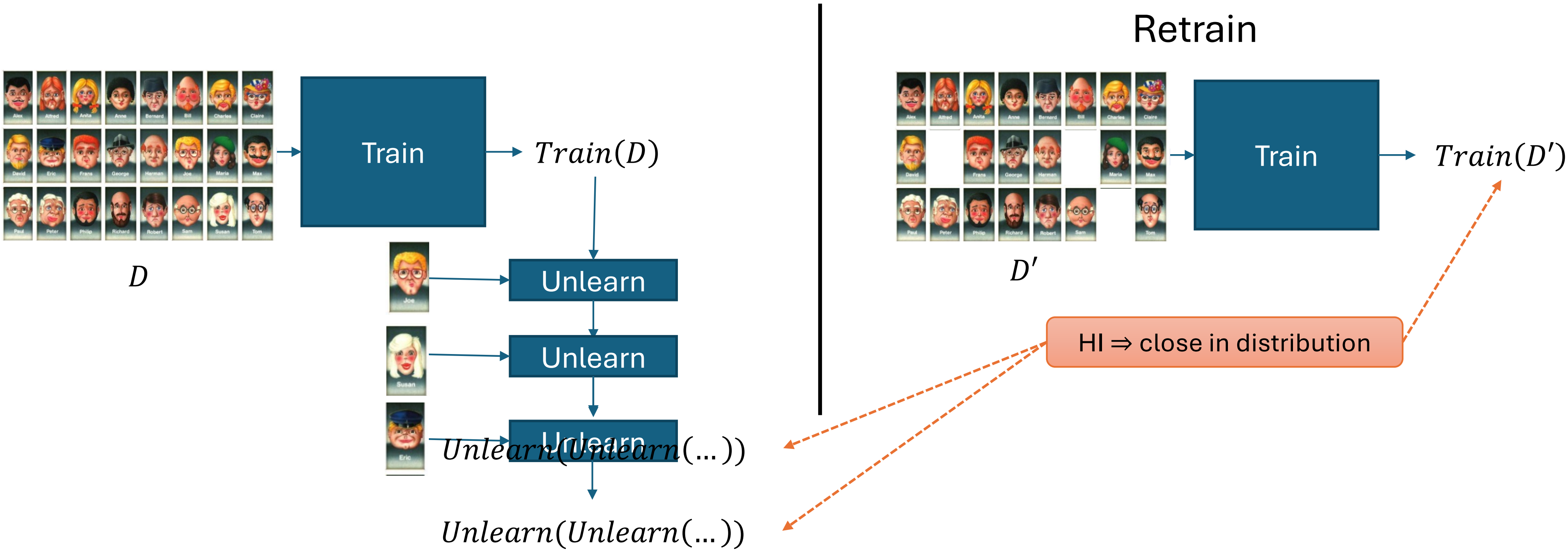
$$M(X) \approx_{\epsilon,\delta} M(X')$$

Let's suppose DP anonymizes.
See: US Census, Facebook, Apple, Google, ...

History-independence vs DP



History-independence vs DP



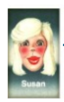
History-independence vs DP

Train

$Train(D)$

D

Unlearn



Unlearn

Unlearn

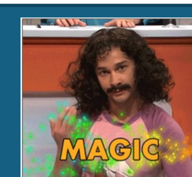
$Unlearn(Unlearn(\ldots))$

Retrain

nothing → Train → nothing

HI ⇒ close in distribution

Differnetial Privacy



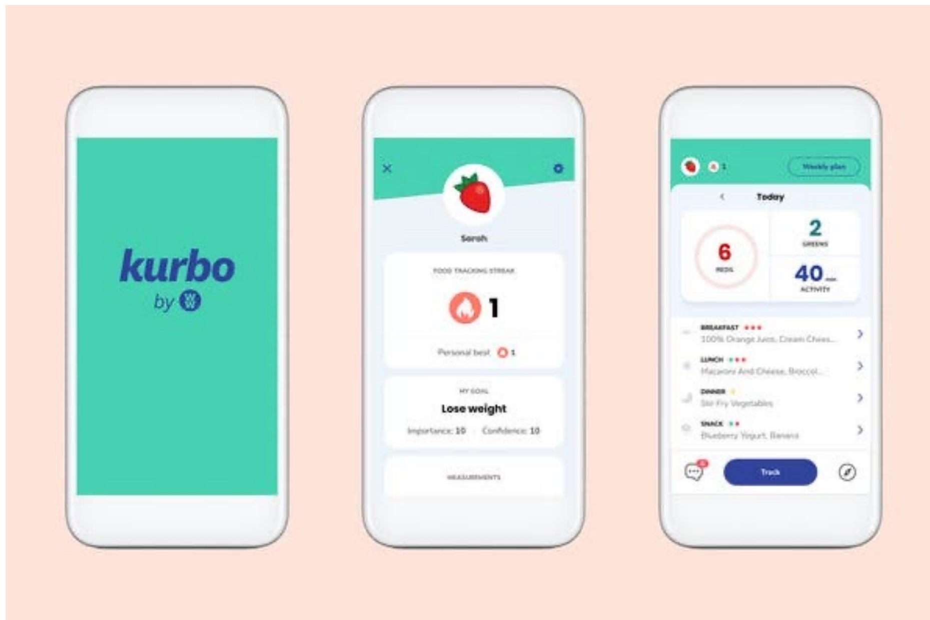
$Train(D)$

Non-trivial DP ML ⇒
DP doesn't satisfy history independence

What does machine unlearning require?

Collective vs individual protection

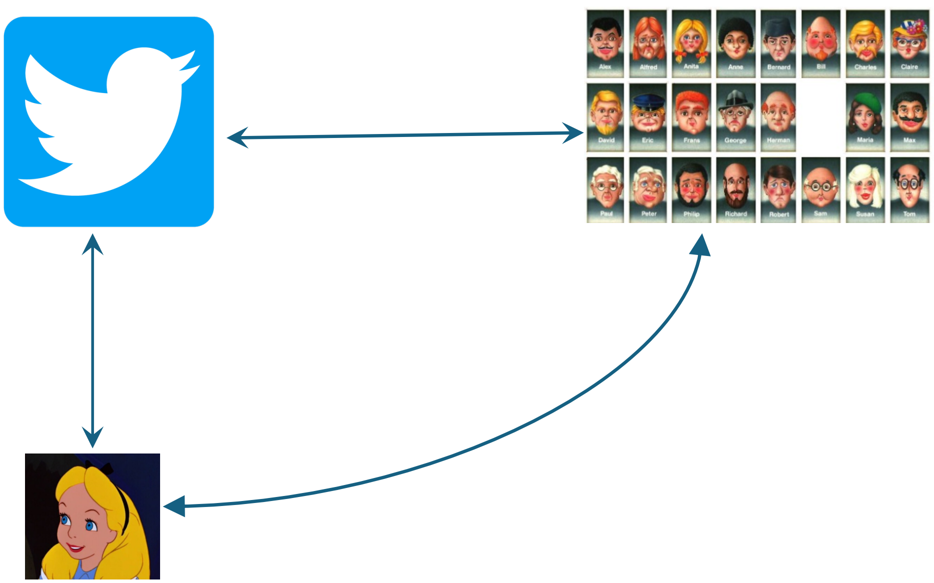
Disgorgement: anonymization is not enough



FTC made WW destroy “any models or algorithms developed in whole or in part using Personal Information Collected from Children through the Kurbo Program”

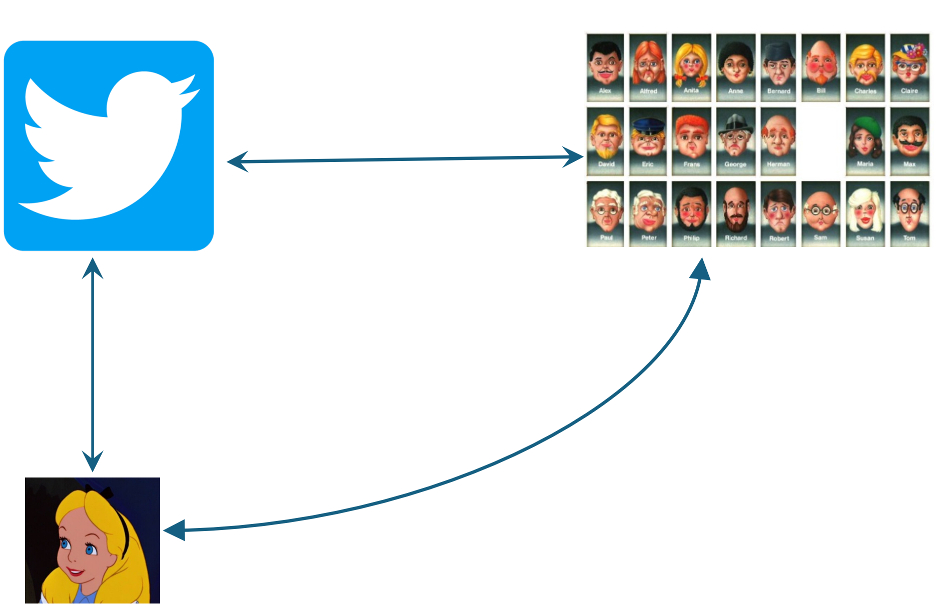
What does data deletion require?

Beyond statistical computations



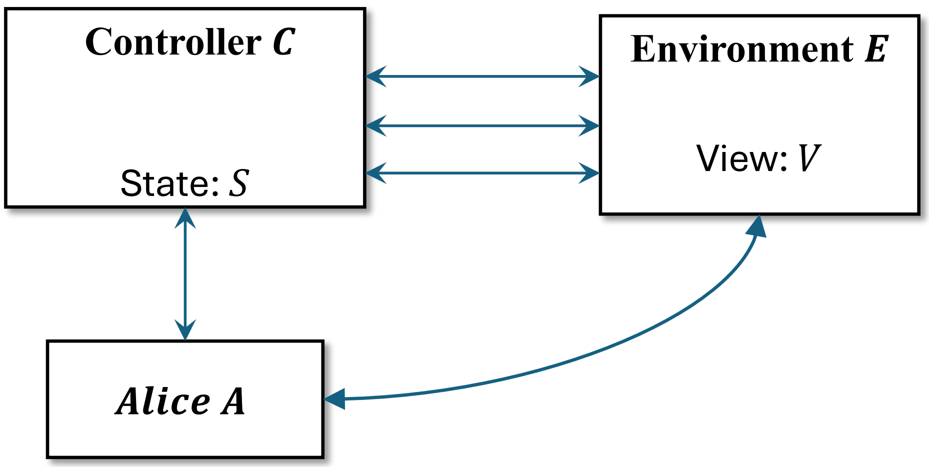
- DP doesn't make any sense for social functionalities
- Can still hope to limit Alice's downstream effect after deletion
- How to formalize?

Beyond statistical computations



- Does anybody think this is **not meaningful**?
- Does anybody think there is a **better** approach?

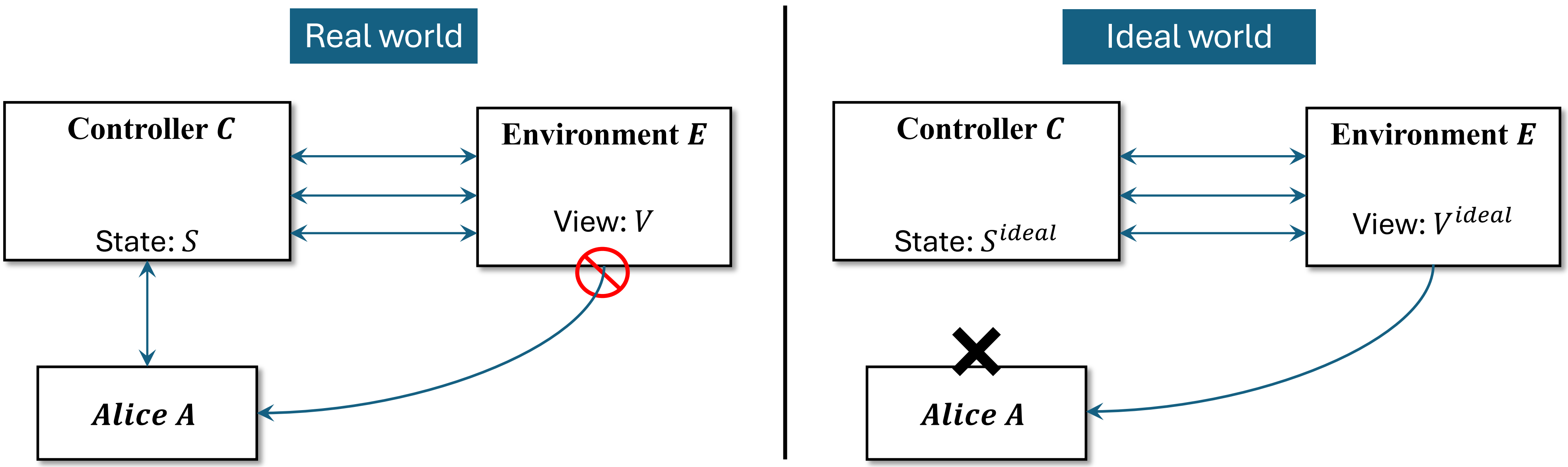
Simplified execution model $\langle C, E, Y \rangle$



- Authenticated channels*
  - One $C \leftrightarrow A$ channel
  - Many $C \leftrightarrow E$ channels
  - C can't distinguish
- Arbitrary interaction
  - Starts with E
  - Send message → activate recipient
  - Ends when A sends DEL to C, and C processes it
- We care about:
  - S: Controller's internal state
  - V: Environment's view

*Authentication is necessary [GGV 20]

Deletion-as-confidentiality [GGV 20]



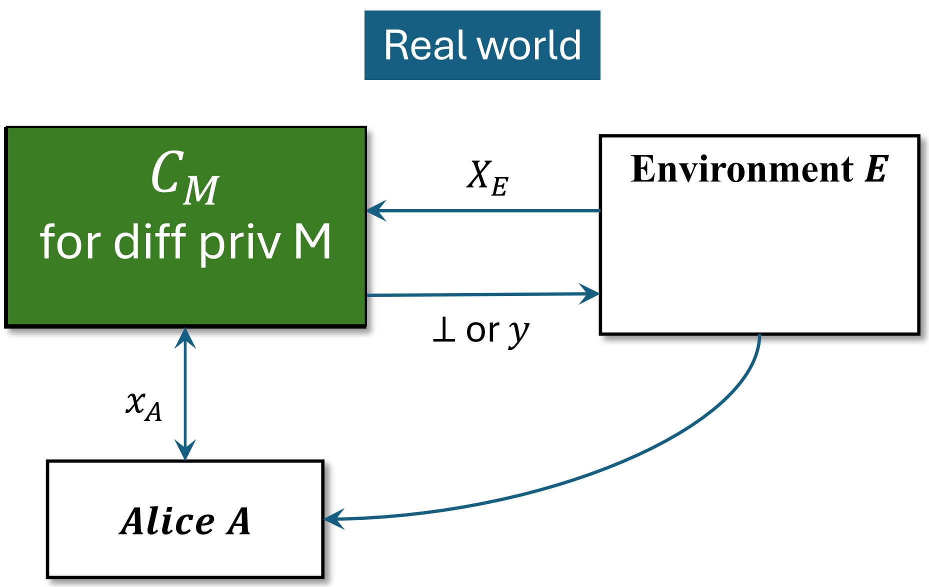
Definition

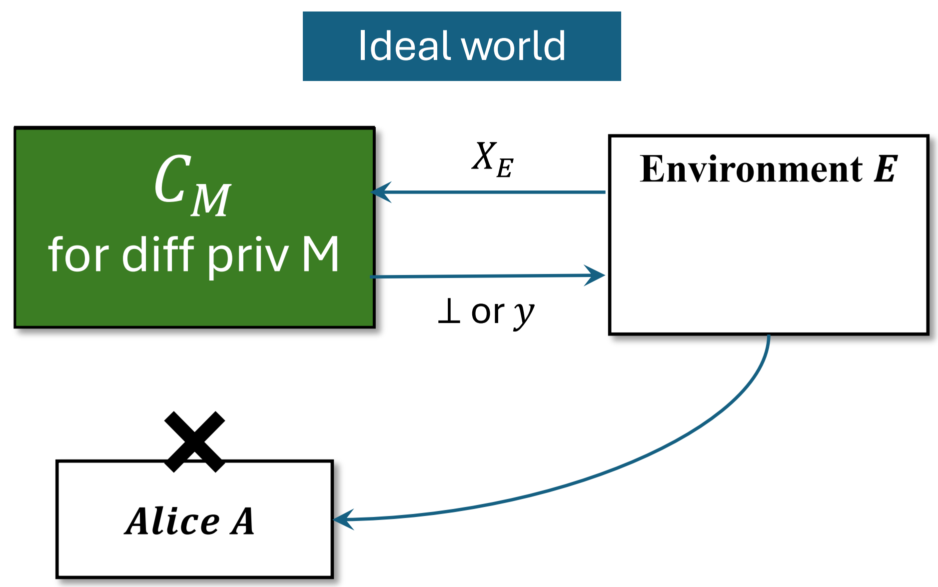
C satisfies (ϵ, δ) – **deletion-as-confidentiality** if for all E, Y

$$(S, V) \quad \approx_{\epsilon,\delta} \quad \left(S^{ideal}, V^{ideal}\right)$$

Adapted from "Formalizing Data Deletion in the Context of the Right to be Forgotten" by Garg, Goldwasser, Vasudevan (2020)

Example: One-shot DP



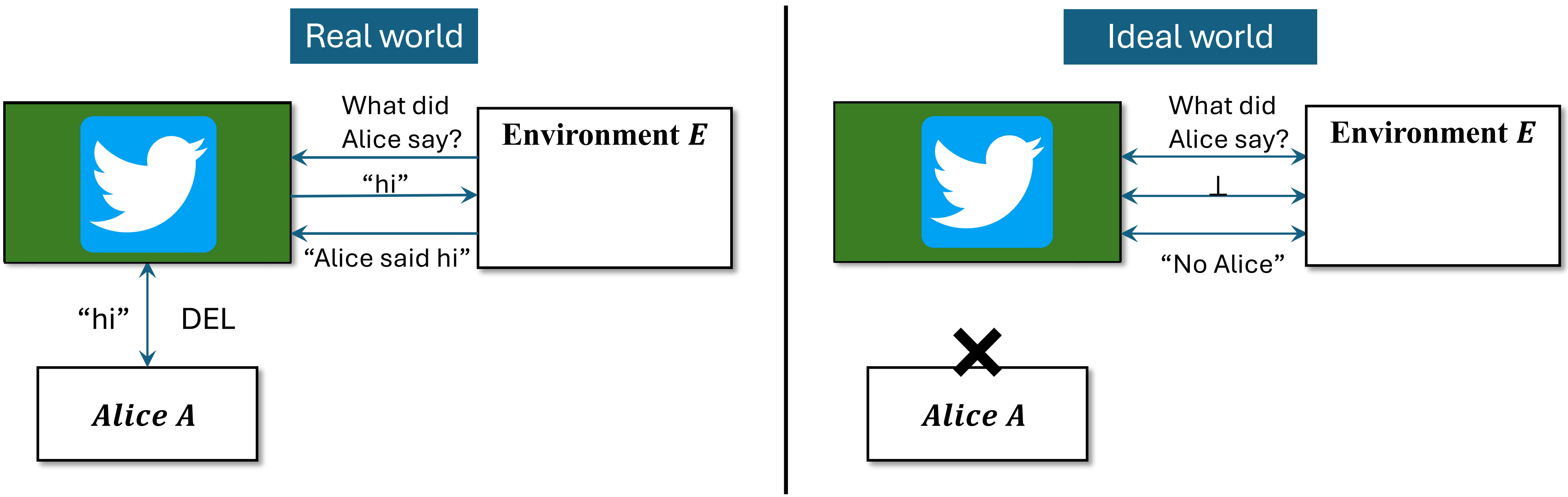


C_M

Before Sept 1, 2024	• Maintain $D = X_E \cup \{x_A\}$ • Return $\perp$
Midnight, Sept 1, 2024	• $y = M(D)$ • Erase D
After Sept 1, 2024	• Return y

- If DEL before Sept 1:
  - $S = X_E = S^{ideal}$ ← need history independence
  - $V = \perp = V^{ideal}$
- If DEL after Sept 1:
  - $S = M(X_E \cup \{x_A\}) \approx_{\epsilon,\delta} M(X_E) = S^{ideal}$
  - $V = M(X_E \cup \{x_A\}) \approx_{\epsilon,\delta} M(X_E) = V^{ideal}$

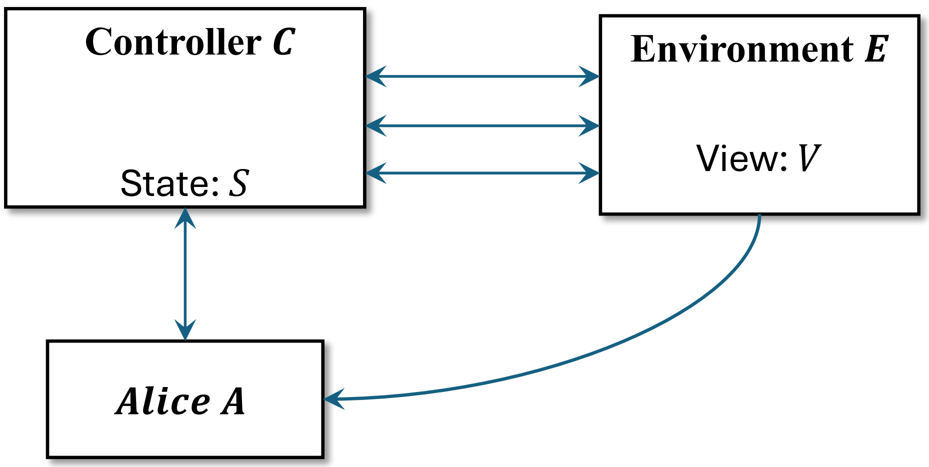
Example: Bulletin Board



Confidentiality ⇒
Alice and Env **never** interact

Confidentiality is too strong:
no bulletin board

Simulatable deletion [GL 22]



Definition

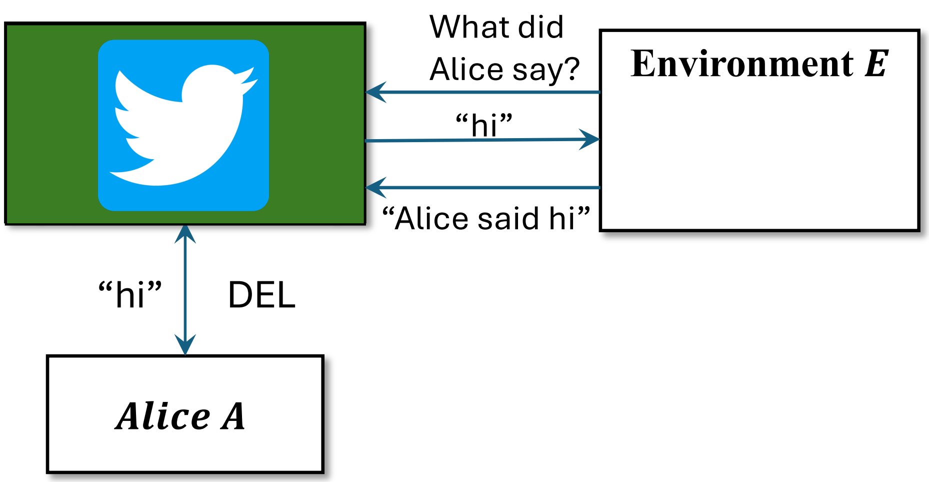
C satisfies **simulatable deletion** if

there exists a simulator Sim such that for all for all E, Y

$$(S, V) \quad \approx \quad (Sim(V), V)$$

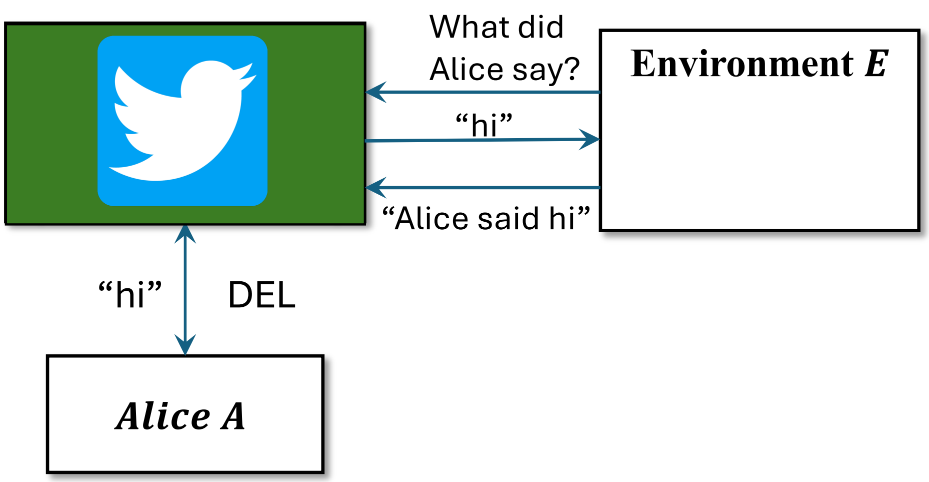
Adapted from "Deletion-Compliance in the Absence of Privacy" by Godin, Lamontagne (2022)

Example: Bulletin Board



- Controller's state: $(\perp,$ "Alice said hi")
- Simulator:
  - Read the transcript
  - Write down all messages from E
- $(S, V) = (Sim(V), V)$ if state history independent

Example: Bulletin Board

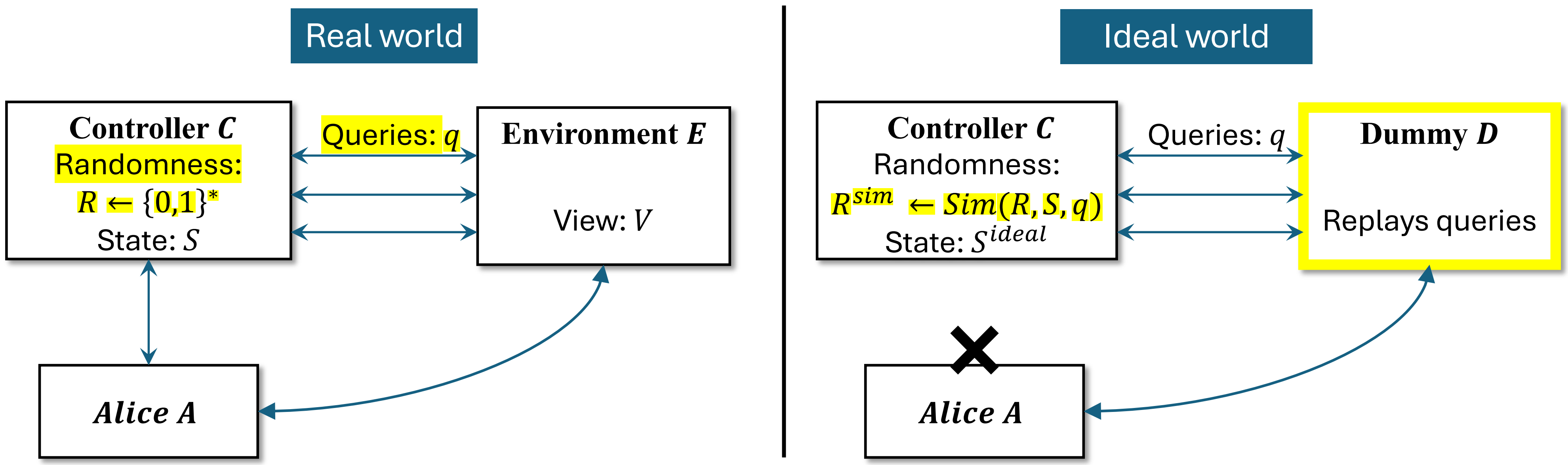


- Controller's state: ("hi" "Alice said hi")
- Simulator:
  - Read the transcript
  - Write down all messages from E and A
- $(S, V) = (Sim(V), V)$ if state history independent

Simulation ⇒
Don't delete anything that was made public

Simulation is too weak:
no deletion!

Deletion-as-control [CSSV 23]



Definition

C satisfies $(\epsilon, \delta) -$ **deletion-as-control** if

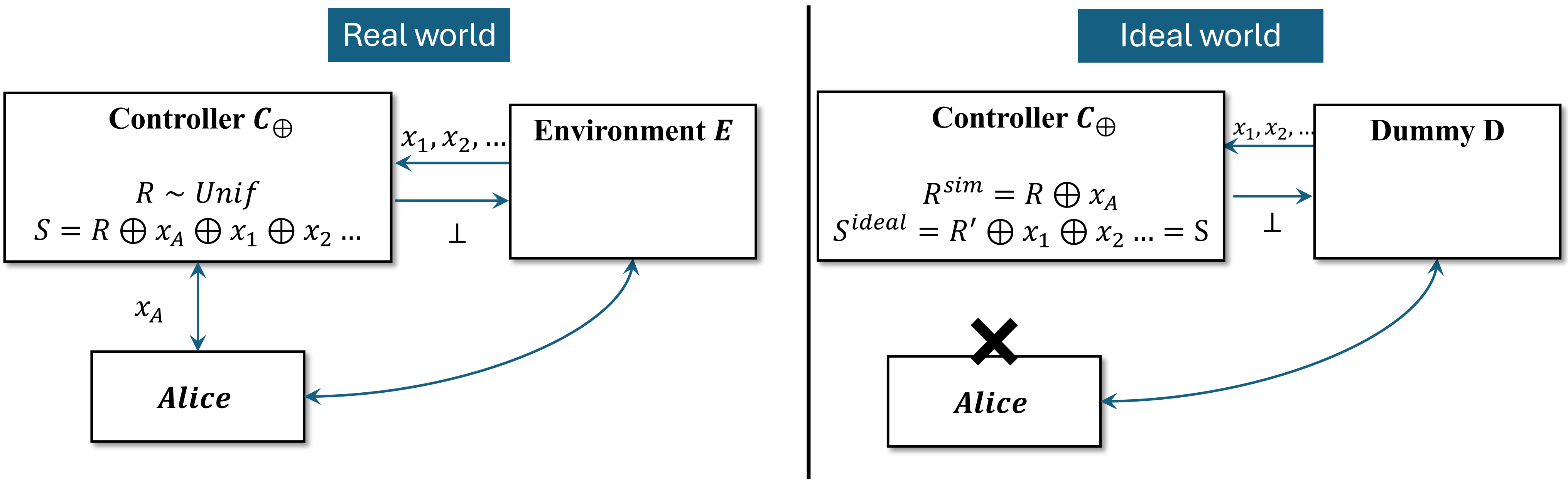
there exists a simulator Sim such that for all for all E, Y

- $\Pr[S^{ideal} = S] \geq 1 - \delta$
- $R^{sim} \approx_{\epsilon,\delta} Unif$

R^{sim} is plausible

Together, q and R^{sim} explain the state S

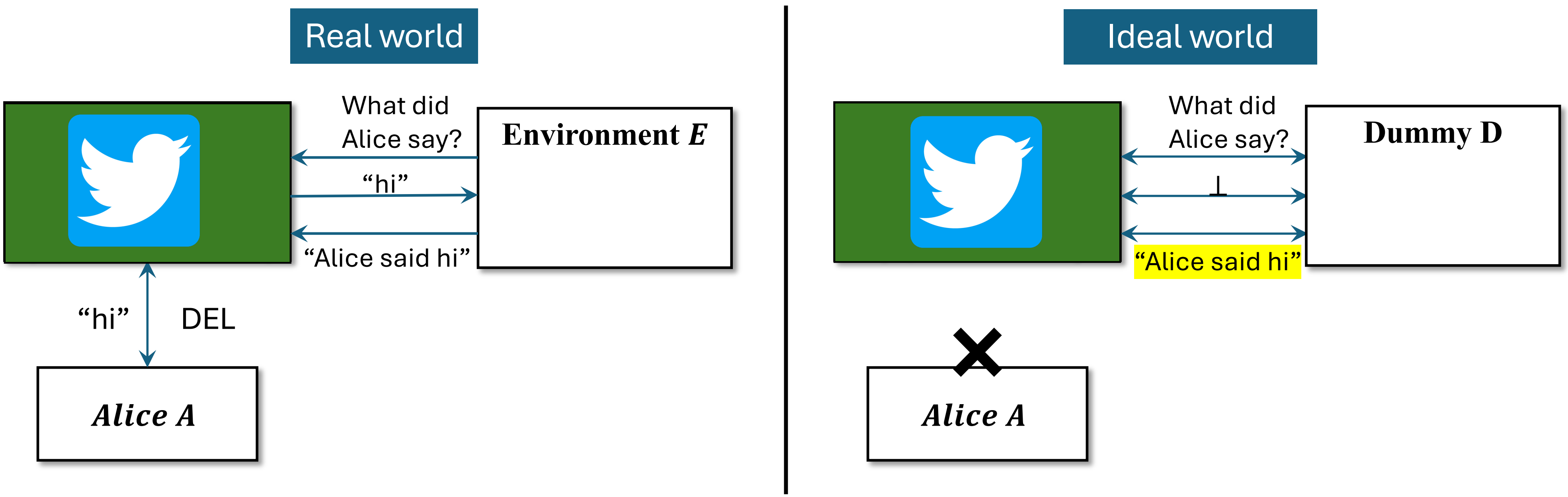
Example: XOR



Claim: $C_\oplus$ satisfies (0,0)-deletion-as-control.

- $\Pr[S^{ideal} = S] = 1$
- $R^{sim} \sim Unif$

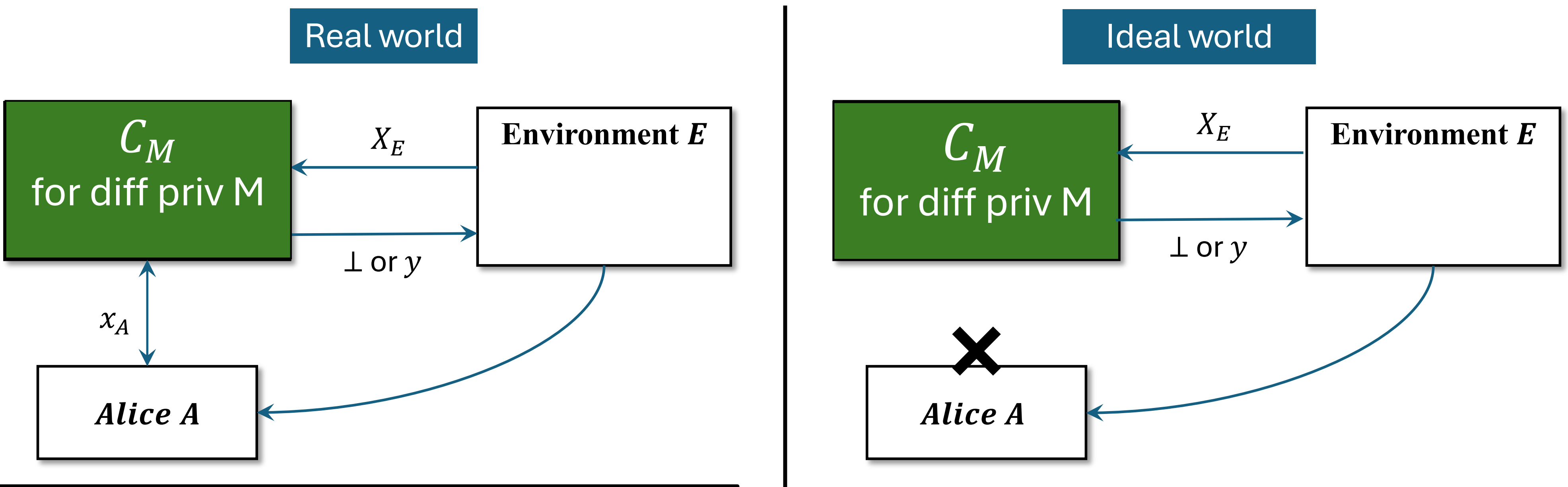
Example: Bulletin Board



In both worlds: Lingering dependence on A iff E’s msgs depend on A’s msgs

Theorem
C is history independent ⇒
(0,0)-deletion-as-control

Example: One-shot DP



Lemma: If $M(D;R) \approx_{\epsilon,\delta} M(D';R)$, then sampling R then R' conditioned on equality gives:

- $\Pr[equal] > 1 - \delta$
- $R' \approx_{\epsilon,\delta} R$

Example:

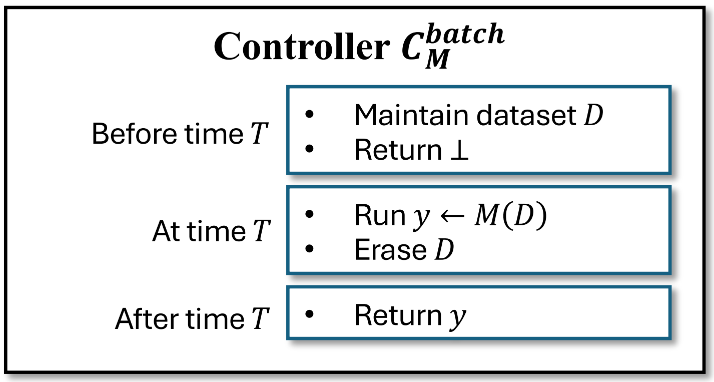
$M(D) = \sum x_i + R$ for $R \sim Lap\left(\frac{1}{\epsilon}\right)$

$R^{sim} = R + x_{alice}$ is (ϵ, δ)-close to $Lap\left(\frac{1}{\epsilon}\right)$

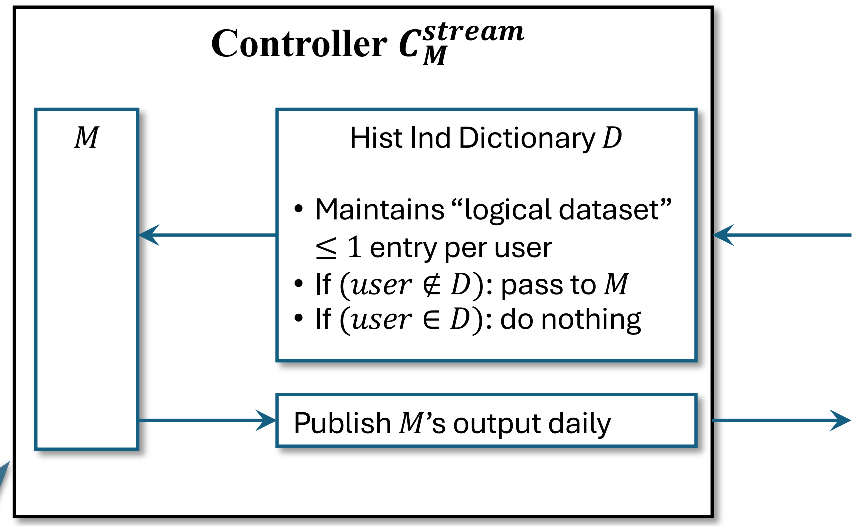
(ϵ, δ) DP $\Rightarrow$ (ϵ, δ) deletion-as-control

Eg: DP-FTRL [KMSTTX 21]

Theorem 1: Batch processing; central DP

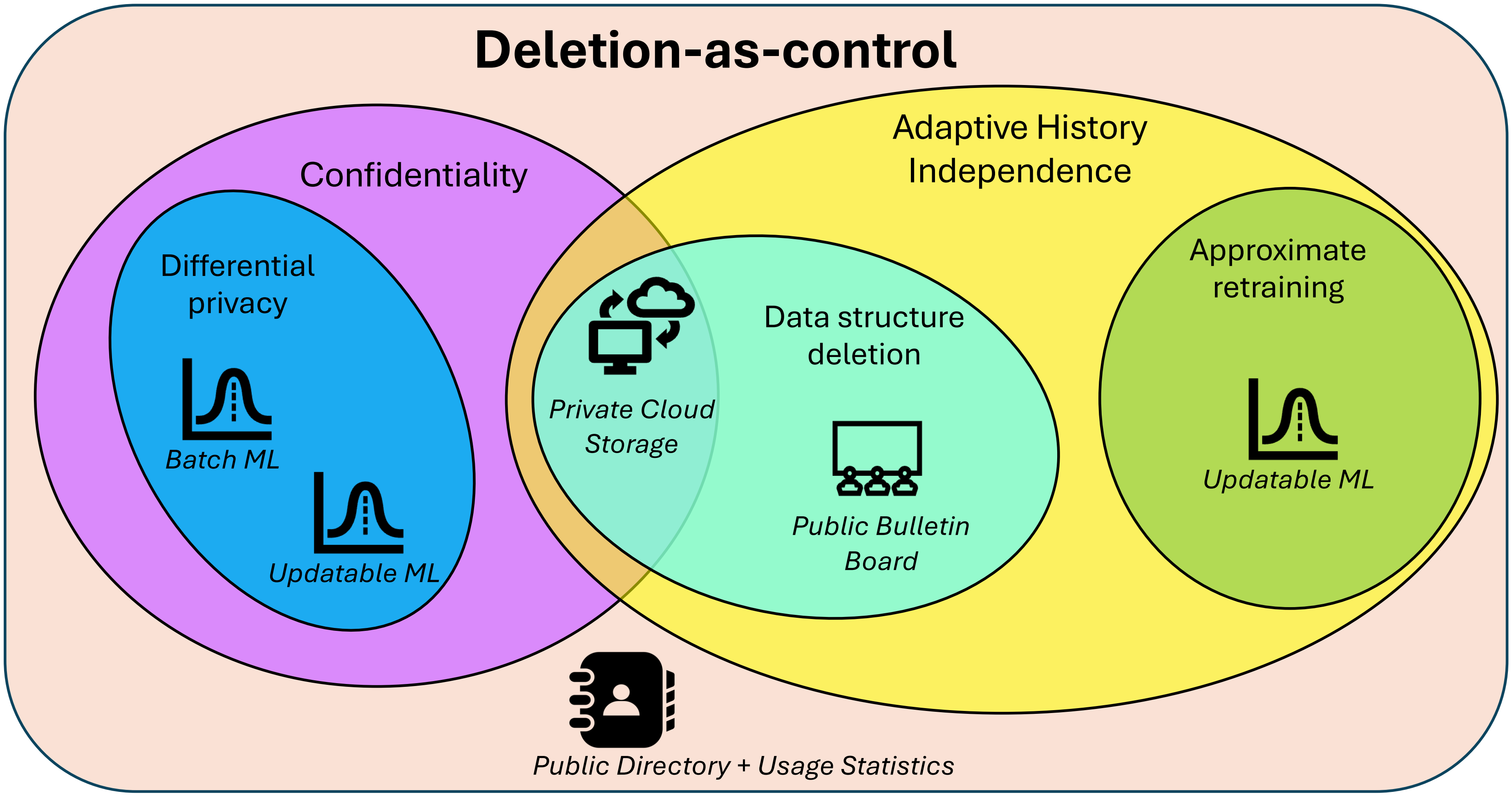


Theorem 2: Streaming processing; event-level, adaptive pan-privacy + continual release



Formally, requires defining:

- Adaptive PP + CR
- Adaptive HI
- Adaptive execution of arbitrary interactive TMs
- Interfaces stitching them all together


Deletion-as-control
Confidentiality
Differential privacy
Batch ML
Updatable ML
Adaptive History Independence
Private Cloud Storage
Data structure deletion
Public Bulletin Board
Approximate retraining
Updatable ML
Public Directory + Usage Statistics

Machine unlearning and anonymization

1. **Extract** relevant text and examples
2. **Formalize** mathematically
3. **Analyze,** alone and in relation to other notions
4. **Draw** legal conclusions

Legal conclusions

- K-anonymity (and related techniques) fail as general purpose anonymizers
- Some support for the view that DP anonymizes. If so…
- New MUL algorithms / tradeoffs possible
- Different contexts → different requirements
  - Collective (disgorgement) vs Individual (erasure) rights


AND NOW FOR SOMETHING COMPLETELY DIFFERENT
Copyright © MCMLXXI Kettledrum/Lownes Productions Ltd.
1 / 10