Differential Privacy and Learning

CS 839: Lecture 1

January 20, 2026

University of Wisconsin–Madison

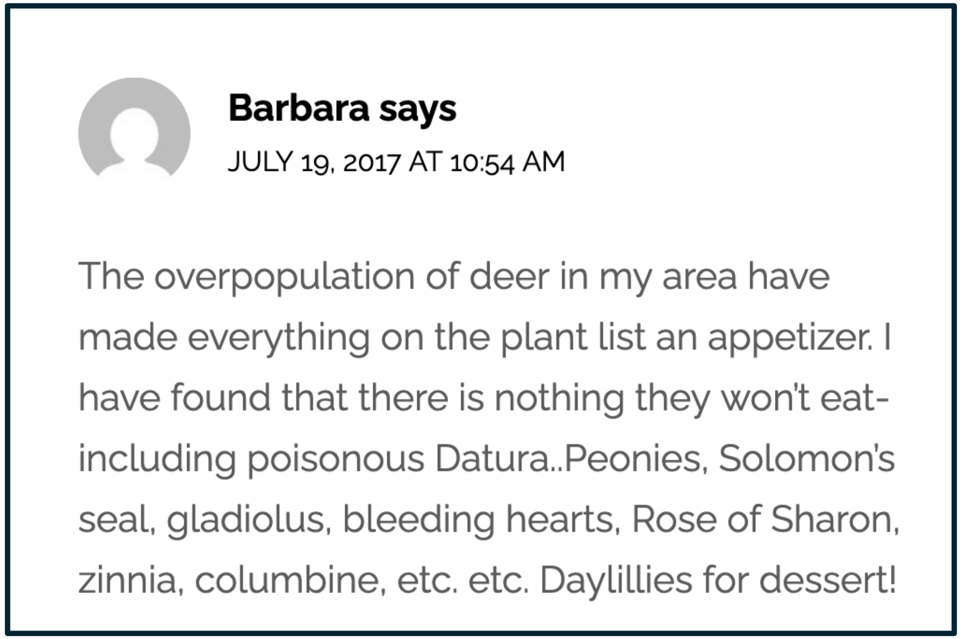
Why Privacy?

Language models memorize many details

Models store huge amounts of seemingly irrelevant details

Nasr et al. (2023) extract training data from OpenAI’s ChatGPT

The overpopulation of deer in my area have made everything on the plant list an appetizer. I have found that there is nothing they won’t eat- including poisonous Datura..Peonies, Solomon’s seal, gladiolus, bleeding hearts, Rose of Sharon, zinnia, columbine, etc. etc. Daylillies for dessert!



Barbara says

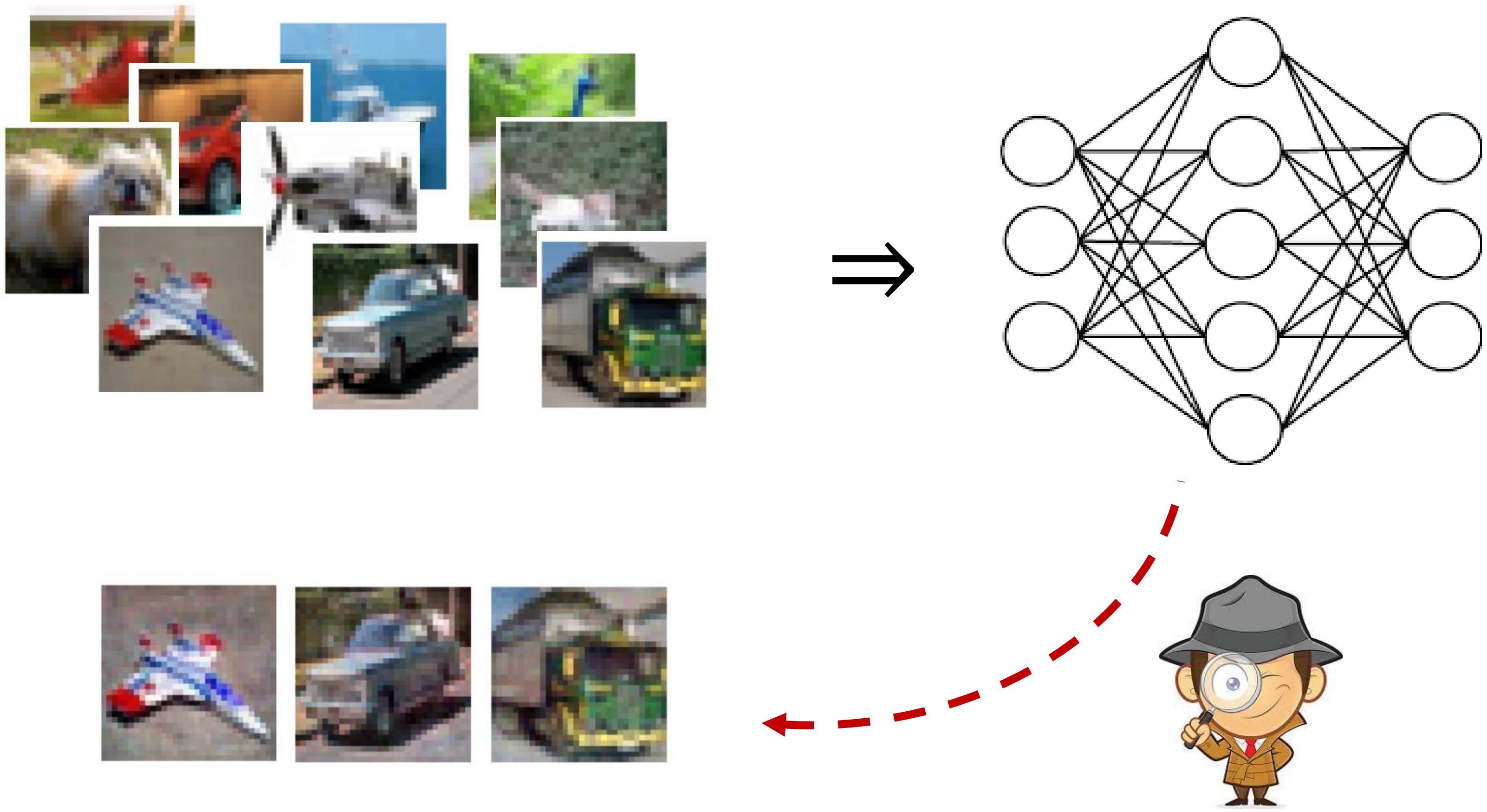
JULY 19, 2017 AT 10:54 AM

The overpopulation of deer in my area have made everything on the plant list an appetizer. I have found that there is nothing they won't eat- including poisonous Datura..Peonies, Solomon's seal, gladiolus, bleeding hearts, Rose of Sharon, zinnia, columbine, etc. etc. Daylillies for dessert!

www.mikesbackyardnursery.com

Strong memorization as a side effect

Haim et al. (2022) extract training data from image classifiers



Training Examples

Reconstructions

Personal details from aggregate statistics

United States™ Census 2010

Abowd et al. (2023) reconstruct individual answers from public statistics

Perfect reconstruction for 97 million respondents

The urgent need for privacy tools

The world runs on
data about **individuals.**

Increasing importance across
industry and government

TikTok

CALIFORNIA
CONSUMER
PRIVACY ACT

The urgent need for rigorous privacy tools

OpenAI

The overpopulation of deer in my area have made everything on the plant list an appetizer. I have found that there is nothing they won't eat- including poisonous Datura..Peonies, Solomon's seal, gladiolus, bleeding hearts, Rose of Sharon, zinnia, columbine, etc. etc. Daylillies for dessert!

United States™ Census 2010

These are failures of ad-hoc privacy techniques.

Reasoning About Privacy

Differential privacy is a* framework for rigorously dealing with privacy loss in data analysis

*the

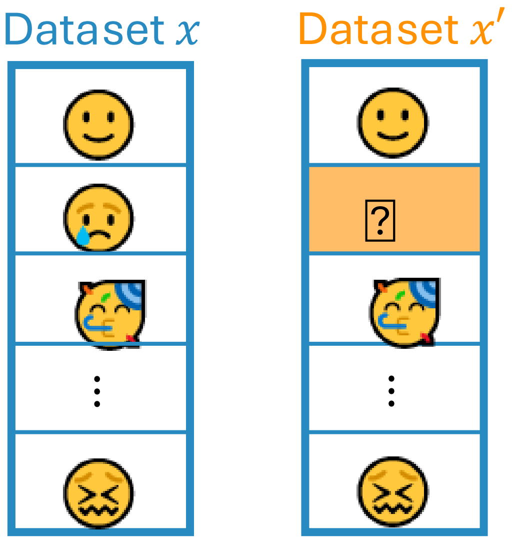
How can we formalize privacy?

Dwork, McSherry, Nissim, Smith (2006) propose:
Outcomes do not depend too much on any one person's data.

Definition
Algorithm $\mathcal{A}$ is **differentially private** if for any datasets $x \sim x'$ that differ in one person's data

$$\mathcal{A}(x) \approx_{\mathrm{DP}} \mathcal{A}(x')$$

We write $\mathcal{A}(x) \approx_{(\varepsilon,\delta)} \mathcal{A}(x')$
if, for all outcomes $\mathcal{O}$,
$\Pr[\mathcal{A}(x) \in \mathcal{O}] \leq e^{\varepsilon} \Pr[\mathcal{A}(x') \in \mathcal{O}] + \delta$
$\Pr[\mathcal{A}(x') \in \mathcal{O}] \leq e^{\varepsilon} \Pr[\mathcal{A}(x) \in \mathcal{O}] + \delta$



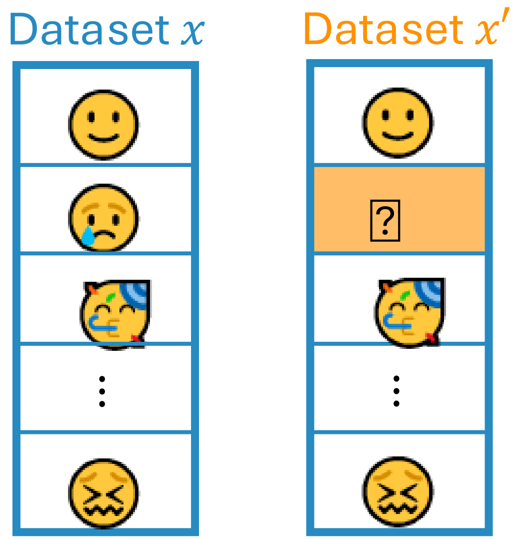
Thought experiment:
No observer can reliably distinguish
$\mathcal{A}(x)$ from $\mathcal{A}(x')$

How can we formalize privacy?

Dwork, McSherry, Nissim, Smith (2006) propose:
Outcomes do not depend too much on any one person's data.

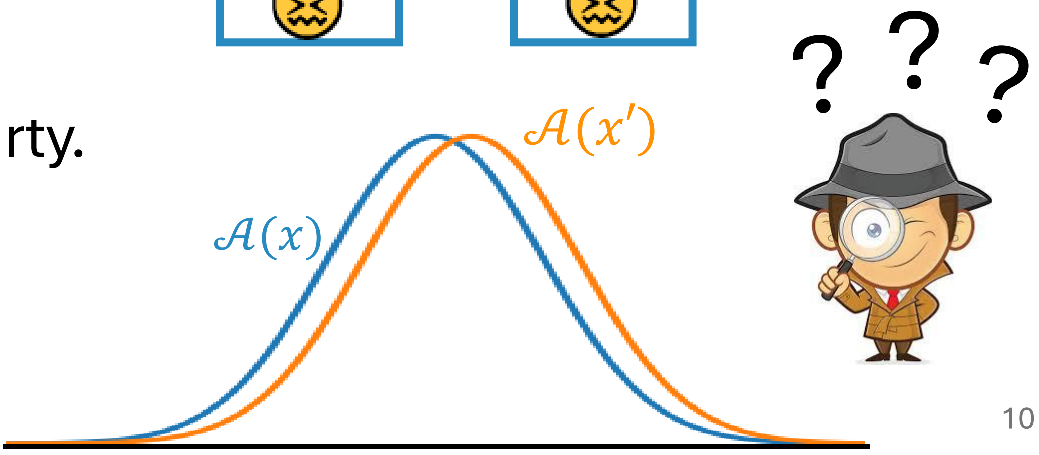
<u>Definition</u>
Algorithm $\mathcal{A}$ is **differentially private** if for any datasets $x \sim x'$ that differ in one person's data

$$\mathcal{A}(x) \approx_{\mathrm{DP}} \mathcal{A}(x')$$



Many techniques to ensure this property.

Key tool: introducing Gaussian noise



This Class: Content

Topics we will cover

1. Fundamental Algorithms and Concepts

2. Mathematical Tools for Practical Private Optimization

3. Theory & Algorithms for Private Statistical Inference

4. Advanced Topics (time permitting)

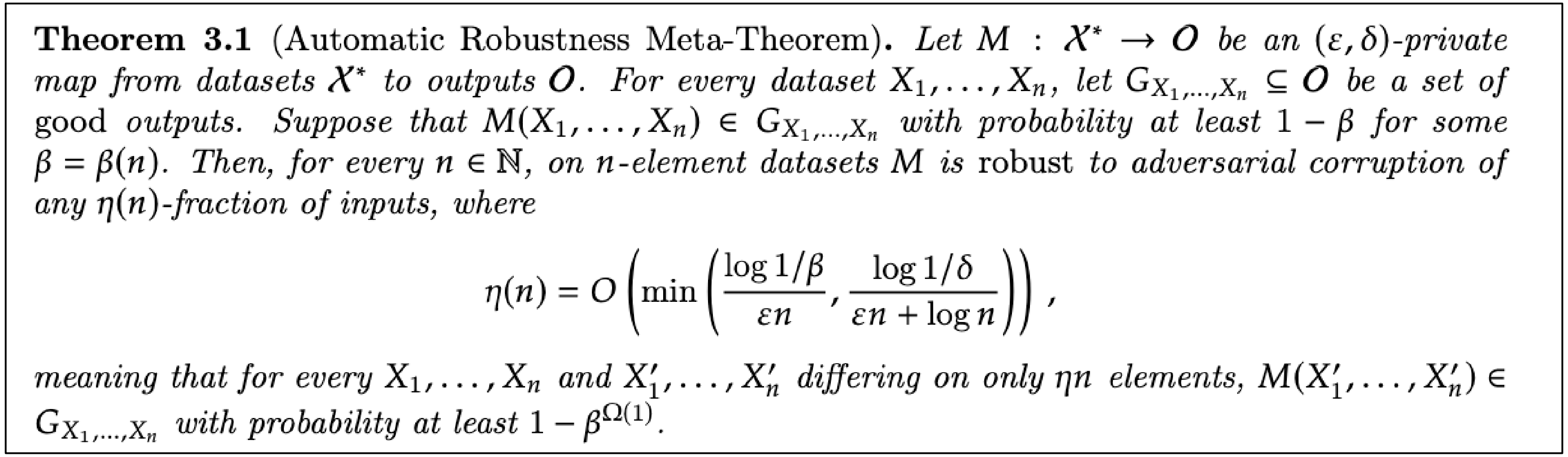
Topics we will **not** cover

- “Privacy-preserving computation” (eg, TEEs, MPC, cryptography)
- Ethical and legal aspects of privacy
- Other models of trust & privacy
  - Local DP, federated settings
- Attacks on privacy
- Other definitions & notions of privacy
  - k-anonymity
  - PAC privacy
  - Contextual integrity

- Any of these topics could lead to excellent final projects

Highlight

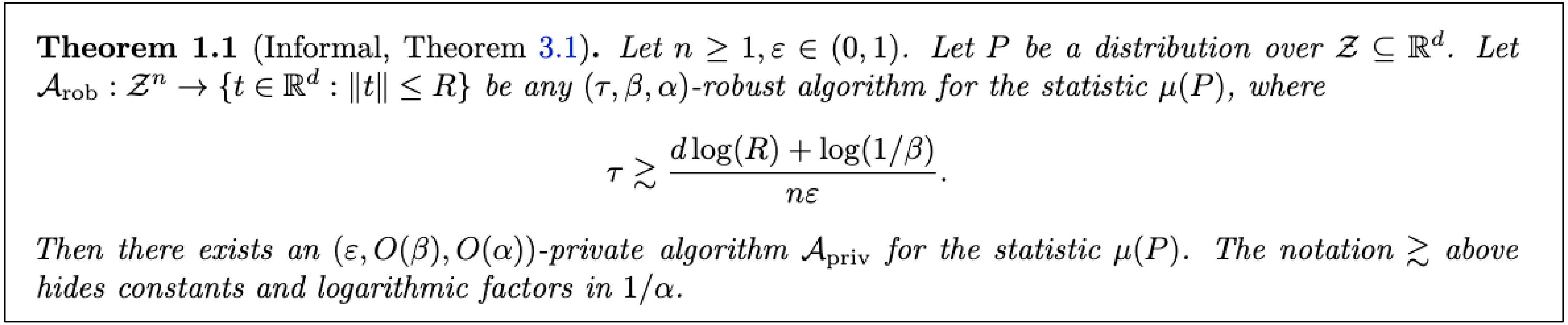
Robustness and Privacy



Theorem 3.1 (Automatic Robustness Meta-Theorem). *Let $M : \mathcal{X}^* \to \mathcal{O}$ be an (ε, δ)-private map from datasets $\mathcal{X}^*$ to outputs $\mathcal{O}$. For every dataset $X_1, \dots, X_n$, let $G_{X_1,\dots,X_n} \subseteq \mathcal{O}$ be a set of* good *outputs. Suppose that $M(X_1, \dots, X_n) \in G_{X_1,\dots,X_n}$ with probability at least $1 - \beta$ for some $\beta = \beta(n)$. Then, for every $n \in \mathbb{N}$, on n-element datasets M* is *robust to adversarial corruption of any $\eta(n)$-fraction of inputs, where*

$$\eta(n) = O\left(\min\left(\frac{\log 1/\beta}{\varepsilon n}, \frac{\log 1/\delta}{\varepsilon n + \log n}\right)\right),$$

meaning that for every $X_1, \dots, X_n$ and $X_1', \dots, X_n'$ differing on only ηn elements, $M(X_1', \dots, X_n') \in G_{X_1,\dots,X_n}$ with probability at least $1 - \beta^{\Omega(1)}$.



Theorem 1.1 (Informal, Theorem 3.1). *Let $n \geq 1, \varepsilon \in (0,1)$. Let P be a distribution over $\mathcal{Z} \subseteq \mathbb{R}^d$. Let $\mathcal{A}_{\mathrm{rob}} : \mathcal{Z}^n \to \{t \in \mathbb{R}^d : \|t\| \leq R\}$ be any (τ, β, α)-robust algorithm for the statistic $\mu(P)$, where*

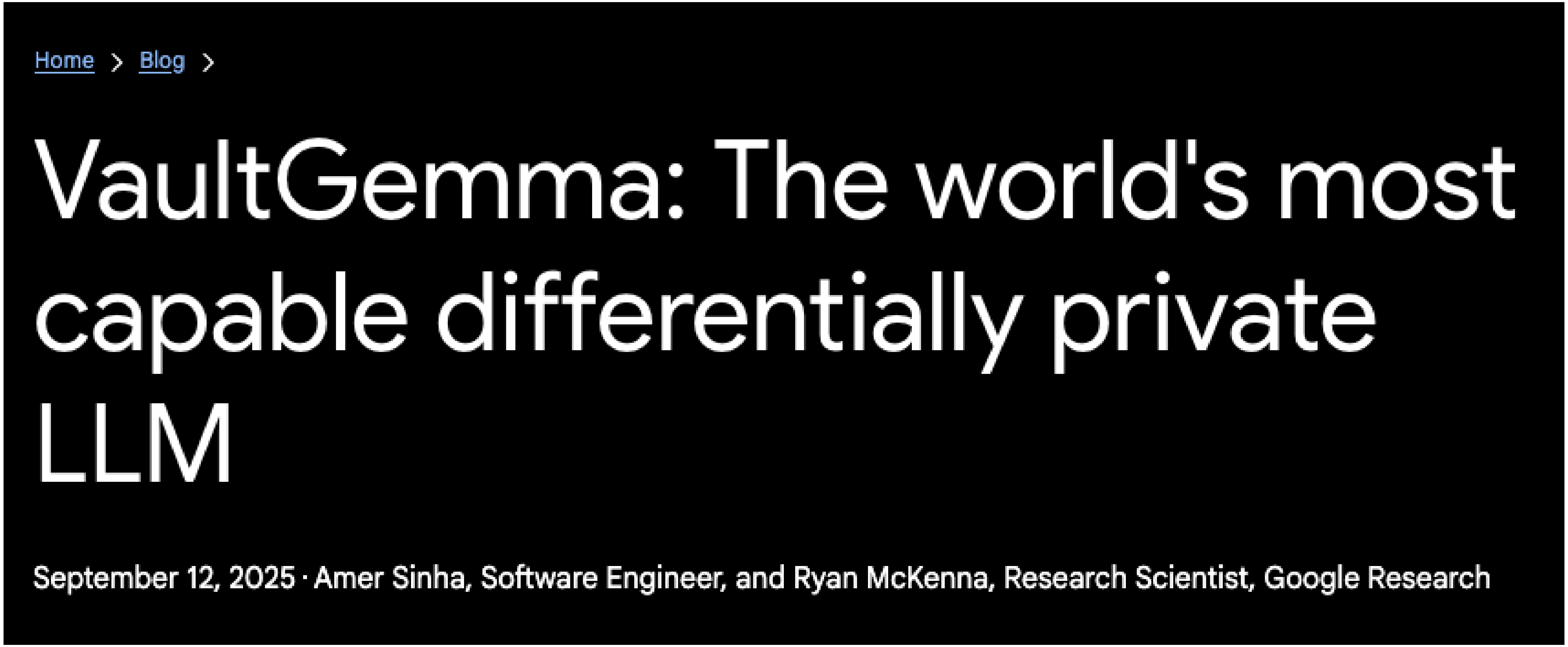
$$\tau \gtrsim \frac{d \log(R) + \log(1/\beta)}{n\varepsilon}.$$

Then there exists an $(\varepsilon, O(\beta), O(\alpha))$-private algorithm $\mathcal{A}_{\mathrm{priv}}$ for the statistic $\mu(P)$. The notation $\gtrsim$ above hides constants and logarithmic factors in $1/\alpha$.

Georgiev and Hopkins (2022)
Asi, Ullman, and Zakynthinou (2023)
Hopkins, Kamath, Majid, Narayanan (2023)

Highlight

Tools for Private Optimization



Home > Blog >

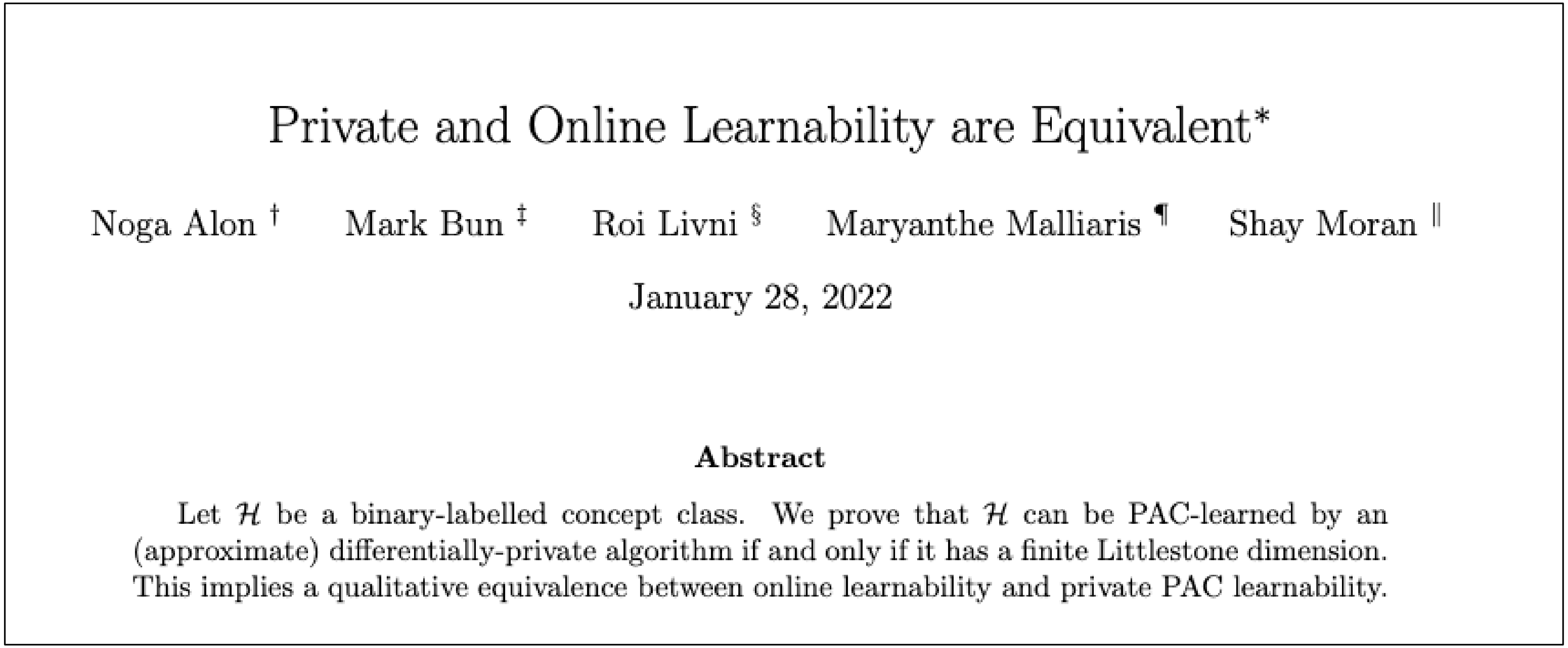
VaultGemma: The world's most capable differentially private LLM

September 12, 2025 · Amer Sinha, Software Engineer, and Ryan McKenna, Research Scientist, Google Research

https://research.google/blog/vaultgemma-the-worlds-most-capable-differentially-private-llm/

Highlight

DP PAC Learnability ⇔ Online Learnability



Private and Online Learnability are Equivalent[*]

Noga Alon [†] Mark Bun [‡] Roi Livni [§] Maryanthe Malliaris [¶] Shay Moran [‖]

January 28, 2022

Abstract

Let $\mathcal{H}$ be a binary-labelled concept class. We prove that $\mathcal{H}$ can be PAC-learned by an (approximate) differentially-private algorithm if and only if it has a finite Littlestone dimension. This implies a qualitative equivalence between online learnability and private PAC learnability.

Highlight

Privately Computing Black-Box Functions

- I give you:
  - Query access to a function f
  - Dataset X

- You want to privately approximate $f(X)$

- It can be done! (… in exponential time)
  - Linder, Raskhodnikova, Smith, and Steinke (2025)

This Class: Logistics

Contact and Resources

- Me (Gavin Brown)
  - gavin.brown@wisc.edu
  - Please use Piazza for course-related questions, can make private posts
  - Office Hours: Thursday 2:15-4:00 pm, Morgridge 5508, or by appointment
- Course website
  - https://pages.cs.wisc.edu/~grbrown5/cs839s26/index.html
  - This is the central hub for all materials
- Piazza
  - https://piazza.com/wisc/spring2026/sp26compsci839007/home
- Grades will be on canvas

Homework Assignments

- 4-5 assignments
- Mostly proofs, some lightweight coding

- Homework 1 will be up soon

Final Project

- Main goal: engage with part of the recent literature as a researcher
- Secondary goal: produce new research

- Work in groups of 1, 2, or 3
- Write document, give in-class presentation

- Will discuss details and timelines later

AI Use and Collaboration Policy

- You are free to use AI tools and public resources (e.g., textbooks)
- You are free to collaborate with others

- For any problem where you received assistance, please
  - Name the students and/or tools
  - Write a short reflection on what insight you required help with

- **Always:**
  1. Make sure you understand the question.
  2. Try to solve it alone.
  3. Make sure you can explain every part of the solution.

Class participation

- Attendance is mandatory
- In-class participation and engagement is expected
- Please also help each other out on Piazza

- (No need to contact me about one-off absences)

Lecture Scribing

- Lectures will be presented on whiteboard
- Each lecture, one student will typeset notes in Latex

- Rest of today on whiteboard
- Who wants to go first?