Scientific Research Publishing

Open Journal of Social Sciences, 2025, 13(12), 399-410
https://www.scirp.org/journal/jss
ISSN Online: 2327-5960
ISSN Print: 2327-5952

Differential Privacy Implementation for Anonymous Student Feedback on Campus Safety and Belonging

Emma Liu[1], Joyce Guo[2]

[1]Kent School, Kent, CT, USA
[2]Booth School of Business, The University of Chicago, Chicago, IL, USA
Email: emmaliu.apps@gmail.com, joyceguo78@gmail.com

How to cite this paper: Liu, E., & Guo, J. (2025). Differential Privacy Implementation for Anonymous Student Feedback on Campus Safety and Belonging. *Open Journal of Social Sciences, 13,* 399-410.
https://doi.org/10.4236/jss.2025.1312030

Received: October 10, 2025
Accepted: December 23, 2025
Published: December 26, 2025

Abstract

This paper is a review of differential privacy in data collection. Differential privacy is a mathematical framework that protects individual privacy when third parties collect and analyze sensitive information. The system works by adding carefully controlled mathematical noise to datasets to conceal any specific person's data in the analysis. We will further explore its current applications in the fields of healthcare and public policy and detail our program developed upon this foundation. Utilizing differential privacy to maintain anonymity, the program is a survey that collects student feedback within a high school or college setting. The goal of this project is to help schools better understand student experiences and concerns while ensuring that personal information remains confidential and protected.

Keywords

Differential Privacy, Data Anonymization, Local Differential Privacy, Student Feedback, Campus Safety, Privacy-Preserving Data Collection, Educational Data Ethics, Laplace Mechanism, Privacy Budget, Anonymous Surveys

1. Introduction

In today's data-driven society, organizations collect huge amounts of personal data to inform decisions and improve services. However, such heightened data usage raises serious concerns about the privacy of individuals. Traditional anonymization methods are widely used: hospitals remove patient names from medical records before sharing data for research, schools strip student names from academic performance data, and companies remove customer identifiers before an-

alyzing purchasing patterns. However, these methods often fail to fully protect identities. This occurs because they fail to account for indirect identifiers, such as age, gender, ZIP code, etc. Such pieces of information, when combined with other publicly available data, can be used to re-identify individuals. The cross-referencing process works by finding unique or rare combinations of characteristics. For instance, there might be only one 65-year-old male living in a specific small town who visited a cardiologist on a particular date, making him easily identifiable even in "anonymous" medical data. Anyone with access to these identifiers could cross-reference to pinpoint an individual, making the data collection insecure. The consequences of such re-identification can be severe: individuals may face institutional discrimination or even legal repercussions based on their private data. For students, this could lead to disciplinary actions. Hence, differential privacy was developed to protect against this kind of risk, ensuring that no individual's data has a significant influence on the output. By "no significant influence", we mean that the statistical results should remain virtually the same whether any particular person's data is included or excluded from that dataset.

There are two main types of differential privacy: local and central differential privacy. In central differential privacy, noise is added by the collector to the entire dataset right before it is released. In local differential privacy, which is the method that our implementation utilizes, noise is added to each individual's data before it is sent to the collector. Local differential privacy provides stronger privacy guarantees since the raw data is protected from the collector. However, it typically requires more noise to achieve the same effectiveness as central differential privacy.

Differential privacy is a mathematical framework that tackles this issue. By inserting carefully calibrated randomness into datasets, differential privacy renders the presence or absence of an individual's data insignificant to the overall outcome. Inserting randomness means adding mathematical noise, small random numbers, to either individual responses or to the final statistical results. E.g., for medical data, you might add small random values to the patient's age or test results. In a survey, you could randomly flip some "yes" answers to "no" with a small probability. The randomness is not arbitrary but calculated to provide strong privacy guarantees while maintaining the usefulness of the data for statistical analysis. Differential privacy requires careful implementation and can reduce data accuracy, especially for small datasets. The process is typically handled by data scientists or privacy engineers who implement the algorithms. While differential privacy provides strong mathematical guarantees, it could theoretically be "broken" if implemented incorrectly or if the privacy parameters are set poorly. Furthermore, differential privacy has significant practical limitations that restrict its applicability. First, it operates on a "privacy budget. This means that each query or analysis consumes some of this budget, and once exhausted, no further queries can be made without compromising privacy guarantees. This means organizations must decide in advance what analyses they want to perform. Second, all potential queries must be known at the outset of data collection, preventing researchers from

asking new questions that arise during the course of a study. These constraints make differential privacy challenging to implement in research environments where questions evolve based on preliminary findings.

Confidentiality in data collection requires systems that cannot reveal anything substantial about any single person, even to a third party that has access to external information. This level of security is especially important in sensitive environments like schools, where students may feel reluctant to give honest feedback. Differential privacy solves this need by offering a strict, quantifiable guarantee of privacy. It allows the students to answer without fear that their responses will be traced back to them.

Many school surveys struggle to balance transparency and privacy because they often ask sensitive questions about bullying, mental health, or substance use. Traditional school climate surveys all face this challenge: students may not provide honest answers in fear that their responses could be traced back to them, potentially leading to disciplinary action or unwanted attention from faculty. This project integrates differential privacy into the world of student feedback. While surveys are increasingly being used within schools to gauge student well-being and improve the learning environment, they often struggle to balance transparency and privacy. Students should trust differential privacy guarantees because they are based on rigorous mathematical proofs rather than policies. The randomness insertion happens automatically through software running on secure school servers or cloud systems, ensuring that the privacy protection is applied consistently. We would like to solve this problem by developing a program that allows students in high school or college to give honest feedback through a differentially private survey. In this way, we aim to allow schools to make informed, data-driven decisions while maintaining the trust and confidentiality of students.

Technical Background

To further understand differential privacy, we need to define some key terms.

1) computation—any mathematical operation performed on data.

2) dataset—a collection of information about individuals, like survey responses.

3) randomized algorithm—a computer program that uses random number generation each time it runs, even with the same input data.

The core of differential privacy lies in the idea of limiting how much information any single data point contributes to the output of a computation.

$$Pr\left[M\left(D\right)\in S\right]\le \mathrm{e}^{\varepsilon}\times Pr\left[M\left(D'\right)\in S\right]$$

This mathematical equation captures the principle of differential privacy protection (Dwork, 2011). M represents a randomized algorithm, D represents our original dataset, and D' represents a "neighboring" dataset that is identical to D except for one person's data being added or removed. S represents any possible set of outputs that our algorithm might produce.

This equation states that the probability (Pr) of a randomized algorithm (M)

when run on dataset (D) producing a result dataset (S) can be at most e^{ε} times larger than the probability of yielding that same result set through a neighboring dataset (D') that differs by one person's data.

Here, epsilon (ε) is the privacy loss parameter, which controls how much privacy protection we get. We cannot simply choose epsilon to be incredibly small because that would require adding so much noise that our results would become meaningless. If we set epsilon to exactly 0, we would need infinite noise, making our data completely unusable. Differential privacy does require randomness. Without it, an attacker could run the same analysis multiple times and average out the noise.

The mechanism most commonly used to achieve this is the Laplace Mechanism, which adds carefully calibrated noise. The amount of noise added depends on the sensitivity of the function and the epsilon value. This mathematical framework provides provable privacy guarantees, meaning we can mathematically demonstrate that privacy will be protected. It also maintains statistical utility, meaning the noisy data can still be used to draw accurate conclusions about overall trends and patterns. However, this doesn't work equally well for all functions. Simple counting queries and basic statistics work well because they have low sensitivity and require minimal noise. Complex analyses may become unreliable because they suffer from noise accumulation and higher sensitivity. For example, an analysis trying to detect relationships between student safety and academic performance might conclude no relationship exists when the noise drowns out the actual correlation.

While pure differential privacy uses the guarantee

$$Pr\left[M\left(D\right)\in S\right]\le e^{\varepsilon}\times Pr\left[M\left(D'\right)\in S\right]$$

practical systems often adopt approximate DP, adding a small failure probability $\delta>0$:

$$Pr\left[M\left(D\right)\in S\right]\le e^{\varepsilon}\times Pr\left[M\left(D'\right)\in S\right]+\delta.$$

This $\left(\varepsilon,\delta\right)$ form facilitates mechanisms like the Gaussian mechanism (often preferred for high-dimensional or composition-heavy workloads). The key design knob is global sensitivity Δf, the maximum change to a function's output when one person's data changes. Mechanisms scale noise to Δf:

- Laplace mechanism (pure DP, previously discussed): add $\mathrm{Lap}\left(0,b\right)$ with $b=\Delta f/\varepsilon$.
- Gaussian mechanism (approx. DP): add $N\left(0,\sigma^2\right)$ with σ proportional to $\Delta f\sqrt{2\ln\left(1.25/\delta\right)}$.
- Exponential mechanism: select categories (e.g., argmax) with DP when outputs are non-numeric.
- Randomized response/k-ary randomized response (local DP): privatize discrete answers by flipping with calibrated probabilities; aggregate with an unbiased estimator.

For means of bounded numeric responses in $\left[L,U\right]$, Δf depends on context:

- Central DP, mean over n users: $\Delta f = (U-L)/n$.
- Local DP, per-response release: if you release a privatized individual value $x \in [L,U]$, the sensitivity of $f(x)=x$ is $\Delta f = U-L$. Scale/clip first to minimize Δf.

Rule of thumb for accuracy: If each of n users releases $x + \mathrm{Lap}(0, \Delta/\varepsilon)$, the sample mean remains unbiased and the added noise's standard deviation is $\sqrt{2}\cdot(\Delta f/\varepsilon)/\sqrt{n}$. A handy back-of-the-envelope for expected absolute error of the mean is $\approx \frac{2}{\sqrt{\pi n}}\cdot\frac{\Delta f}{\varepsilon}$.

Every released statistic spends part of the privacy budget. With basic composition, ε (and δ) add across queries. For a 10-question survey, you can:

1) Allocate a uniform budget (e.g., total $\varepsilon_{tot} = 2 \Rightarrow \varepsilon = 0.2$ per item), or,

2) Use utility-weighted budgeting, granting larger ε to items the school deems critical for decision-making (e.g., safety) and making smaller ε elsewhere.

Advanced composition (and privacy accountants) can slightly improve cumulative guarantees, but the high-level takeaway holds: plan your analysis up front and track spending so repeated re-queries don't silently degrade protections.

2. Applications

Now that we have established the technical foundation, we can examine the current implementations of differential privacy.

2.1. Healthcare

One example of differential privacy's application in healthcare is its use in the diagnosis of coronary heart disease. In Ficek et al. (2021), one study is described that developed a differentially private algorithm specifically for "diagnosing coronary heart disease using medical records" (p. 2273). The algorithm allowed personal health data, such as data collected through smartphones or smartwatches, to be disrupted with noise before leaving the device. This preserved privacy by not letting individual-level data be disclosed in its raw form. Precisely, the algorithm maintained diagnostic accuracy rates above 85%, which is considered clinically useful for screening purposes. However, widespread clinical adoption of such differential privacy systems remains limited, as most are still in research phases due to regulatory concerns. The main drawbacks include reduced accuracy for rare conditions and difficulty integrating with existing medical record systems. Despite the privacy-preserving noise, the algorithm maintained strong performance at "predictive modeling", showing that one can build effective diagnostic tools without compromising patient privacy. The authors also acknowledge, however, that a broad challenge of differential privacy for medical research is that "diminished accuracy in small datasets is problematic" (p. 2269). This shows both the potential and challenge of differential privacy in real clinical practice.

2.2. Public Policy

Another example of differential privacy in use is within the U.S. Census Bureau

for public policy. According to Feldman (2020), the Census Bureau began applying differential privacy to protect individuals in publicly released demographic statistics. Traditionally, census microdata was anonymized by removing names and direct identifiers, but researchers later discovered that combining datasets could still lead to reidentification. The paper explains that differential privacy provides "provable privacy protection against a wide range of potential attacks" (p. 3) by adding noise to outputs, ensuring that the inclusion or exclusion of any one person has little effect on the result, as stated previously. This prevents privacy breaches even when datasets are queried repeatedly. However, the authors also note challenges: "providing information about small or sparse subpopulations is hard to do while providing strong privacy guarantees" (p. 5). Social scientists have raised concerns that this could limit studies on "poverty, inequality, immigration, internal migration, and more" (p. 8). Still, the case of the Census demonstrates the seriousness with which policymakers are beginning to adopt differential privacy to enhance transparency and confidentiality in national statistics.

From our case studies, we can observe the following key points:

1) Differential privacy works best with large datasets where individual noise has less impact on overall patterns.

2) The technique requires careful parameter tuning to balance privacy and utility.

3) Institutional adoption faces practical challenges beyond the technical implementation (Cummings et al., 2024).

4) The approach may inadvertently harm research on marginalized communities who most need policy attention.

These insights inform our approach to student surveys, where we must consider both the benefits of honest feedback and the potential drawbacks of reduced accuracy for smaller student subgroups.

3. Methods

3.1. Threat Model and Re-Identification Pathways

A rigorous privacy analysis begins with an explicit threat model. We assume an adversary who (i) can observe released aggregates or noisy records, (ii) may hold auxiliary data (e.g., social media, public records, or institutional rosters), and (iii) can issue or infer multiple statistics over time. Re-identification typically exploits (a) linkage attacks, where quasi-identifiers (age, ZIP code, time of event) are matched across datasets, (b) differencing attacks, which subtract near-identical aggregates to infer a single record's contribution, and (c) composition attacks, where multiple releases gradually erode privacy guarantees. Differential privacy directly addresses (b) and (c) by bounding how much any single record can influence any output; careful product and platform design (e.g., rate limiting, access control) addresses (a).

3.2. Algorithm and Code Implementation

Our differential privacy survey system follows this algorithm design:

1) Initialization: Set up survey questions and privacy parameters (epsilon, upper and lower bounds, etc.).

2) Initial Response Collection: Collect respondent's answer to each question.

3) Noise Addition: Apply the Laplace mechanism to add calibrated noise to each response.

4) Data Storage: Store the noisy responses.

In this program, the Laplace mechanism is used within a custom Python class to collect and privatize responses. Each survey answer is recorded and then randomized using Laplace noise before being stored. By incorporating this noise at the point of data collection, the system enforces local differential privacy, ensuring that privacy is preserved even before the data is aggregated or transmitted. This approach is particularly valuable in environments like schools, where trust and anonymity are essential for honest feedback.

```
from diffprivlib.mechanisms import LaplaceTruncated
import numpy as np

#survey, applying diff priv
class DiffPrivSurvey:
    def __init__(self, questions, epsilons=None, lower=1, upper=10):
        self.questions = questions
        # Allow per-question ε; default to uniform if not provided
        self.epsilons = epsilons or= [10] * len(questions)
        self.lower = lower
        self.upper = upper
        self.all_responses = []
    def _clamp(self, v):
        return max(self.lower, min(self.upper, v))

    def collect_response(self, question_idx, question):
        response = int(input(f"{question} ({self.lower}-{self.upper}): "))
        x = self._clamp(response)
        delta = self.upper - self.lower
        mech = LaplaceTruncated(epsilon = self.epsilons[question_idx], sensitivity=delta, lower=self.lower, upper=self.upper)
        noisy_response = mech.randomise(x)
        return round(noisy_response)

    def run_survey(self):
        for i, q in enumerate(self.questions):
            response = self.collect_response(i, q)
            self.all_responses.append(response)

    def get_responses(self):
```

```
            return self.all_responses

        def print_responses(self):
            print(self.all_responses)

questions = [
        "I feel safe and secure while on campus.",
        "I feel like I belong at this school.",
        "I feel comfortable expressing my identity at school.",
        "Bullying or harassment is handled fairly by the school.",
        "I feel stressed or overwhelmed by schoolwork.",
        "I am provided with adequate academic support.",
        "I am provided with adequate mental health resources.",
        "I trust that the school takes student feedback seriously.",
        "Students at this school are treated equally regardless of race, gender, or back-
ground.",
        "I feel motivated to do well academically at this school.",
]

survey = DiffPrivSurvey(questions)
survey.run_survey()
survey.print_responses()

I feel safe and secure while on campus. (1-10):    8
I feel like I belong at this school. (1-10):    5
I feel comfortable expressing my identity at school. (1-10):    9
Bullying or harassment is handled fairly by the school. (1-10):    2
I feel stressed or overwhelmed by schoolwork. (1-10):    4
I am provided with adequate academic support. (1-10):    6
I am provided with adequate mental health resources. (1-10):    7
I trust that the school takes student feedback seriously. (1-10):    1
Students at this school are treated equally regardless of race, gender, or back-
ground. (1-10):    4
I feel motivated to do well academically at this school. (1-10):    9

[7, 5, 6, 3, 3, 6, 7, 3, 5, 8]
```

The DiffPrivSurvey class is designed to collect sensitive student feedback while preserving individual privacy through differential privacy techniques. It defines a list of Likert-scale questions on students' experiences and perceptions of their school environment, ranging from feelings of safety to access to mental health resources. When the survey is run, each question is presented to the respondent, who enters a numerical answer from 1 to 10. Instead of storing the raw response directly, the system adds mathematically calibrated noise to each answer to ensure

that individual data points cannot be precisely traced back to any participant. This allows for meaningful data aggregation while protecting personal information.

For each response, the Laplace mechanism, implemented via the IBM diffprivlib library (Holohan, 2025), adds noise using a randomization process defined by the user's privacy budget, epsilon. The mechanism introduces noise in proportion to the sensitivity of the function being computed. The sensitivity equals 9 because we're measuring how much a single person's response can change the output. As the survey is on a scale of 1 to 10, the maximum difference is bounded by 9 units.

The privacy loss parameter ε controls the balance between privacy protection and data accuracy in differential privacy. Smaller ε values provide stronger privacy guarantees but inject more noise, reducing utility; larger ε values weaken privacy but yield more accurate results.

In practice, ε should be determined by the intended analytical accuracy, the number of individuals contributing, and any additional safeguards in the system (such as local perturbation, encryption, or limited data access).

For local differential privacy systems such as ours, the effect of ε interacts strongly with the sample size n. Because each participant adds noise to their own response, the noise in the aggregated mean decreases roughly as $1/\sqrt{n}$. The approximate standard error due to Laplace noise for bounded values in $[L,U]$ with sensitivity $\Delta = U - L$ is:

$$S_{DP} \approx \frac{\sqrt{2}\Delta}{\varepsilon\sqrt{n}}.$$

Thus, as n grows, even modest ε values produce reliable estimates; conversely, with small n, higher ε is needed to maintain usable signal quality. **Table 1** illustrates the expected accuracy trade-off for a 1 - 10 Likert scale ($\Delta = 9$) at different ε and n values.

Table 1. Expected accuracy (in Likert points) for different sample sizes and privacy parameters

n	$\varepsilon = 0.5$	$\varepsilon = 1$	$\varepsilon = 2$	$\varepsilon = 5$	$\varepsilon = 10$
10	4.0	2.8	1.4	0.6	0.4
50	1.8	1.3	0.6	0.3	0.18
200	0.9	0.6	0.3	0.13	0.09

(Values are approximate S_{DP} in Likert points.)

This relationship highlights that ε cannot be evaluated in isolation. For large surveys (hundreds of participants), ε values between 0.5 and 2 often provide strong privacy with minimal loss of accuracy.

However, in smaller studies, or classroom-level pilots, $\varepsilon = 10$ can be reasonable: it keeps added noise below half a point on a 10-point scale while still randomizing individual answers enough to discourage re-identification.

Moreover, because our system enforces local perturbation—each student's response is privatized before submission—the effective privacy risk is already re-

duced compared with central models. Within this context, $\varepsilon = 10$ offers a practical, transparent trade-off between privacy and interpretability, allowing schools to make meaningful use of aggregated feedback.

Each noisy response is generated and stored in self.all_responses, a list that represents all modified answers for a single survey run. The run_survey() method iterates over each question and collects responses with added noise, and print_responses() outputs the final result as a list of values. These values are privacy-preserving approximations of the original responses. Importantly, although individual responses are obscured, aggregated patterns across many responses can still be meaningfully analyzed, which is the central benefit of applying differential privacy in surveys. For example, the average scores across questions will converge to the true values as sample size increases (since the random noise will cancel itself out). This implementation ensures that even if survey data is accessed or analyzed later, the privacy of any one individual remains protected.

Alternatively, the survey can treat each Likert item as a distinct categorical label, applying k-ary randomized response (kRR). Here, each participant reports their true category with probability

$$p = \frac{e^{\varepsilon}}{e^{\varepsilon} + k - 1}$$

and reports one of the other $k-1$ categories uniformly at random otherwise. This method ensures that all outputs remain valid Likert choices and eliminates the risk of producing out-of-range values. It is especially suitable for estimating proportions or histograms across categories (e.g., "What fraction of students strongly agree?"), rather than for numeric averages.

Both techniques provide the same formal privacy guarantees but make different utility trade-offs. The numeric approach retains higher accuracy for aggregate averages, while the categorical approach offers cleaner interpretability and stronger discrete protection for individual answers.

3.3. Data Collection and Submission Protocol

Although the current DiffPrivSurvey implementation does not yet include cryptographic token management or duplicate-response detection, these mechanisms are conceptually part of the broader system design. In a production deployment, each participant could receive a single-use, anonymous token or blind-signed credential before responding. The server would then verify the token and mark it as used, ensuring one response per participant without storing personally identifiable information. Such a mechanism prevents duplicate submissions and mitigates risks of data inflation or manipulation while preserving anonymity (Greenstadt & Miers, 2015).

3.4. Handling Missing and Partial Responses

Our prototype assumes full participation, but real surveys often include nonresponse. Drechsler and Bailie (2024) show that imputing missing values or apply-

ing nonresponse weights can increase sensitivity and complicate differential privacy guarantees. To avoid this, our system assigns a separate privacy budget ε_i per question: only answered items consume budget, and unanswered ones simply reduce the effective n for that question. We do not perform imputation inside the DP mechanism, preventing additional sensitivity. Aggregate outputs report each n and adjust confidence intervals accordingly, making the trade-off between privacy and statistical power transparent.

4. Discussion

4.1. Aggregation & Accuracy Reporting

To maintain transparency, publish utility diagnostics alongside results:

- Sampling variance vs. privacy variance. Report the standard error of each mean as $\sqrt{\frac{s^2}{n}+\frac{2\Delta^2}{n\varepsilon^2}}$, where s^2 is the empirical variance of the (noisy) responses, n is the sample size, and the second term approximates the additional DP noise for local Laplace.
- Confidence intervals. Use the combined variance for CIs so stakeholders see how uncertainty shrinks as n grows.
- Release policy. Withhold subgroup statistics unless n exceeds a threshold to mitigate outlier influence and protect small cohorts.

4.2. Platform Hardening Beyond DP

When implementing a privacy-preserving system, developers must address potential issues beyond the scope of differential privacy itself. IP addresses must be hidden by VPNs or proxy servers, and timestamps can be randomized or delayed to prevent timing-based correlation attacks. In addition, the survey infrastructure must be secured by encrypted links, secure hosting environments, and data retention policies. A production-grade deployment should also include:

- Client integrity: prevent multiple submissions (e.g., blind-signed, rate-limited tokens) without tracking identities; resist replay by expiring tokens.
- Transport & storage: TLS in transit, KMS-backed encryption at rest, and strict key rotation.
- Metadata minimization: strip or coarsen device, network, and timing fields; use upload jitter to reduce timing correlation.
- Access control & logging: role-based access with short-lived credentials; log data access (on aggregates, not raw inputs).
- Query governance: a privacy accountant to enforce total ε; dashboards that show "budget spent" per survey wave.
- Red-team simulations: regularly test linkage and differencing attacks using synthetic adversarial datasets.

Only through this integration can educational institutions achieve the privacy protection needed to promote candid student feedback.

4.3. Ethical and Equity Considerations

- Even with strong privacy, differential privacy can reduce signals for small or marginalized subgroups. To avoid silencing these voices, couple DP with:
- Minimum-n thresholds and multi-wave aggregation (pool across time) to raise sample size without increasing ε.
- Decision safeguards: when subgroup estimates are too noisy, defer high-stakes decisions or collect additional data with consent and higher ε explicitly communicated to participants.
- Transparent communication: publish plain-language summaries explaining how privacy noise works and what uncertainty means for policy choices.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Cummings, R., Desfontaines, D., Evans, D., Geambasu, R., Huang, Y., Jagielski, M. et al. (2024). Advancing Differential Privacy: Where We Are Now and Future Directions for Real-World Deployment. *Harvard Data Science Review, 6,* 1-123. https://doi.org/10.1162/99608f92.d3197524

Drechsler, J., & Bailie, J. (2024). *The Complexities of Differential Privacy for Survey Data*. https://arxiv.org/abs/2408.07006

Dwork, C. (2011). The Promise of Differential Privacy: A Tutorial on Algorithmic Techniques. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science* (pp. 1-2). IEEE. https://doi.org/10.1109/focs.2011.88

Feldman, V. (2020). *Differential Privacy: Issues for Policymakers*. Simons Institute. https://simons.berkeley.edu/news/differential-privacy-issues-policymakers

Ficek, J., Wang, W., Chen, H., Dagne, G., & Daley, E. (2021). Differential Privacy in Health Research: A Scoping Review. *Journal of the American Medical Informatics Association, 28,* 2269-2276. https://doi.org/10.1093/jamia/ocab135

Greenstadt, R., & Miers, I. (2015). *ANONIZE: A Large-Scale Anonymous Survey System*. Johns Hopkins University Department of Computer Science. https://www.infoq.com/articles/anonize-large-scale-anonymous-survey-system/

Holohan, N. (2025). Differential Privacy Library. *GitHub*. https://github.com/IBM/differential-privacy-library