Balancing Privacy and Utility: Evaluating Distributional Shifts and Accuracy in Differentially Private Synthetic Breached Data

Gabriel Arquelau Pimenta Rodrigues[1], Matheus Noschang de Oliveira[1], André Luiz Marques Serrano[1], Evanei Gomes dos Santos[1], Geraldo Pereira Rocha Filho[2], Edna Dias Canedo[3], Fábio Lúcio Lopes de Mendonça[1], Daniel Alves da Silva[1], Vinícius Pereira Gonçalves[1]

[1]Electrical Engineering Department (ENE) – University of Brasilia, Brasília, DF, Brazil
[2] Department of Exact and Technological Sciences (DCET) – State University of Southwest Bahia, Vitória da Conquista, BA, Brazil
[3] Department of Computer Science (CIC) – University of Brasilia, Brasília, DF, Brazil

{gabriel.arquelau,matheus.oliveira,fabio.mendonca,daniel.alves}@redes.unb.br, {andrelms,ednacanedo,vpgvinicius}@unb.br, geraldo.rocha@uesb.edu.br

Abstract. *With increasing reliance on data-driven technologies, it is fundamental to ensure the privacy of individuals in datasets. This paper investigates the distributional shift introduced by differential privacy using a synthetically generated dataset simulating leaked personal information. We apply the Laplace mechanism to a hypothetical hotel booking adversarial disclosure scenario and analyze the impact of varying privacy budget (ε) and sensitivity (Δf) parameters across 6,120 combinations. Through Jensen-Shannon Distance and Mean Absolute Percentage Error metrics, we quantify distributional shifts and accuracy degradation. Our findings reveal that attributes with higher entropy experience greater shift under noise addition, contributing with parameter tuning strategies for protecting sensitive data while preserving its analytical value.*

1. Introduction

In the information age, with vast amounts of personal data being collected, processed and shared, it is essential to protect the subjects' privacy to prevent misuse. Without robust privacy protections, sensitive information could be exploited, leading, for example, to threats to personal safety, identity theft and frauds. With respect to this, Privacy Risk Assessments help evaluate the risks associated to storing and handling personal information [Wairimu et al. 2024]. This concern is especially relevant when considering emerging technologies, such as Artificial Intelligence (AI) and the Internet of Things, as they may use or gather a significant amount of personal information in sensitive contexts, such as smart cities, healthcare and financial sectors data [Yao et al. 2023]. This also raises compliance issues, as data protection regulations may require the implementation of privacy protection controls [Prokhorenkov 2022].

With the evolution of AI in sensitive sectors [Bohr and Memarzadeh 2020], an additional challenge refers to privacy attacks in machine learning, in which adversaries

exploit vulnerabilities in models to extract sensitive information from training data. Techniques such as membership inference attacks can determine whether a particular data point was used during training, while model inversion attacks can reconstruct private attributes from model outputs [Rigaki and Garcia 2023]. To safeguard this data, Differential Privacy (DP), which consists of mathematical mechanisms to ensure that the inclusion or exclusion of any individual's data has a minimal impact on query results, may be applied in scenarios where data must be analyzed while preserving individuals' privacy. One possible application is in statistical publishing, such as national censuses, with government agencies using DP to release aggregate population statistics without exposing individual records. The U.S. Census Bureau, for instance, has been using differential privacy in its published data since 2020 [Kenny et al. 2021].

Differential privacy protects data by injecting calibrated noise into query responses, so that even with access to the noisy output, an attacker cannot confidently infer individual data points. Different DP mechanisms add noise in different manners, such as using Laplace or Gaussian distributions, and also vary the noise amount as according to inputted parameters, such as the privacy budget (ε), which bounds how much the inclusion or exclusion of a single individual can affect the probability of any output; and the sensitivity (Δf), which is the maximum change in a function's output when one individual's data is modified. The process, however, generates a distributional data shift, which is the change of statistical properties of the input data. When used in machine learning modeling, for example, this mismatch is relevant, as the models rely on the assumption that training and test data are independent and identically distributed [Malinin et al. 2022]. For this reason, DP results in reduced accuracy in the query response, as excessive noise can obscure data patterns, thus rendering the query response private, but useless. Conversely, too little noise will make the output accurate, but possibly more vulnerable to privacy threats. Hence, it is important to find the right balance in the DP parameters, to maintain data usefulness while safeguarding its privacy [Sato and Minamide 2025].

1.1. Contributions and limitations

This work uses a synthetic breached dataset [Sharma and Bantan 2025] to apply Laplace mechanism DP, evaluating the distributional shifts of attributes counts and the deterioration of precision as a function of the variation of the noise parameters, namely the privacy budget (ε) and the sensitivity (Δf). Therefore, the contributions of this work rely on quantifying the privacy-utility tradeoff in simulated adversarial disclosure scenarios, as measured by distributional shifts and accuracy degradation. Also, it provides knowledge on parameter selection when applying DP to different commonly stored features with varying entropy, balancing disclosure risk against preserved utility.

As limitation of the study, it evaluates exclusively the Laplace mechanism in one specific dataset. Different mechanisms, such as Gaussian and Exponential, may behave differently. Furthermore, we do not consider other privacy-preserving techniques, such as k-anonymity, l-diversity, β-likeness and t-closeness.

1.2. Structure of the work

The remainder of this paper is structured as follows. Section 2 provides the theoretical background and presents a comparative analysis of related work, discussing key similarities, differences and research gaps addressed in this study. Section 3 describes the

methodology and materials used, including a dataset description and the DP assessment approach. Section 4 presents and discusses the experimental results, analyzing the implications of the proposed approach. Ultimately, Section 5 summarizes the main findings and suggests future research.

2. Literature review

Privacy is a fundamental right that supports personal security, autonomy and individual freedom. It encompasses the ability of individuals to control access to their personal information, thereby protecting them from risks such as identity theft, intrusive surveillance and potential physical or psychological harm [Canedo et al. 2023]. Recognizing the importance of privacy in digital environments, [Garcia and Ueyama 2024] proposed a blockchain-based solution that enables privacy-preserving and auditable control over sensitive data. Their approach empowers data owners, such as patients in healthcare scenarios, to manage access to their personal information, reducing the likelihood of unauthorized exposure.

Beyond the healthcare domain, energy consumption and generation data are also considered sensitive, as they can reveal detailed behavioral patterns and daily routines of individuals [Rodrigues et al. 2025]. This sensitivity arises from the increasing reliance of smart grids and smart city infrastructures on fine-grained energy data to optimize energy distribution, demand forecasting and resource efficiency. However, such data collection also introduces significant privacy risks, as unauthorized access or misuse may lead to profiling, surveillance, or other forms of intrusion. In response to these concerns, [Lei et al. 2024] proposed PrivGrid, a privacy-preserving system that enables the collection and forecasting of energy-related data while safeguarding user privacy through secure computation and data protection mechanisms.

Still within the context of energy management and smart grids, differential privacy has emerged as a potential solution to mitigate risks associated with the disclosure of sensitive information. In this regard, [Wen et al. 2022] propose FedDetect, a privacy-preserving federated learning framework designed for energy theft detection. The framework combines local differential privacy with homomorphic encryption, enabling secure and decentralized analysis of smart grid data. Notably, FedDetect achieves this level of privacy protection whilst maintaining high accuracy in identifying anomalous energy usage patterns.

Theoretical research combined differential privacy with rate-distortion theory, showing that differentially private mechanisms can be understood from the perspective of loss relative to utility constraints [Mir 2012]. The work demonstrates that many DP mechanisms naturally emerge from entropy-maximizing formulations under privacy constraints. This formalization situates differential privacy within classical information theory and reinforces its role as a principled approach to balancing information disclosure and utility.

In the context of time-varying data, such as infectious disease statistics, preserving privacy while maintaining utility remains a significant challenge. To address the issue of cumulative error in repeated differentially private releases, recent work has proposed the use of Jensen-Shannon Divergence (JSD) as a criterion for selectively applying privacy-preserving mechanisms only when meaningful data changes occur [Cai et al. 2022]. This

approach reduces the absolute and relative errors in the published statistics, offering a balance between privacy protection and data availability in longitudinal data scenarios.

Expanding its applicability beyond energy and healthcare systems, differential privacy has been extensively studied in a variety of domains [Ouadrhiri and Abdelhadi 2022]. As an example, the U.S. Census Bureau announced the adoption of differential privacy for the publication of federal statistics. In response, [Asquith et al. 2022] conducted an evaluation comparing the original 1940 U.S. Census dataset with three differentially private versions, using privacy budget values (ε) of 0.25, 1.0 and 8.0. Their analysis revealed that query accuracy is inversely proportional to the privacy parameter ε, indicating the trade-off between privacy and utility. In our study, we extend this analysis with the investigation of the effects of distributional shifts under a broader range of conditions. We evaluate 6,120 distinct differential privacy configurations, generated by varying 120 privacy budget values across 51 sensitivity levels. This exploration enables a more granular understanding of how privacy and data sensitivity interact in differential privacy implementations.

The U.S. Census Bureau disclosed the use of $\varepsilon = 17.14$ for person-level data and $\varepsilon = 2.47$ for housing unit data[1]. These privacy loss parameters, although significantly higher than those examined by [Asquith et al. 2022] and in our study, were still sufficient to produce notable disparities in the estimated growth rates of certain ethnic groups [Mueller and Santos-Lozada 2022]. This outcome emphasizes the critical importance of carefully selecting differential privacy parameters, as inappropriate configurations can severely compromise data utility and lead to misinformed research decisions.

A variety of open-source differential privacy tools are currently available and, in addition to evaluating the impact of DP parameters, it is also important to consider the quality and performance of the software libraries that implement these techniques. In this context, [Zhang et al. 2023] conducted a comparative study of several DP libraries, focusing on the degradation of data utility as a function of different privacy budget values. Moreover, these tools have also been evaluated in terms of their usability for data practitioners [Ngong et al. 2024].

In contrast, this study adopts a single differential privacy library, namely `diffprivlib`, as the foundation for all experiments. This creates a consistent environment to conduct a focused analysis of distributional shifts resulting from variations in differential privacy parameters. With this implementation, this work aims to provide knowledge regarding the interaction between sensitivity levels and privacy budget values, minimizing variability introduced by differences in library design or internal mechanisms.

Beyond differential privacy, other anonymization techniques, such as k-anonymity, k-map, δ-disclosure privacy and δ-presence, are also valuable for protecting sensitive data. These techniques are implemented in tools such as ARX Data Anonymization and Amnesia, which have been evaluated by [Tomás et al. 2022] in terms of execution time, usability and the quality of anonymization results. Their comparative study provides practical guidance for selecting the most suitable anonymization tool depending on the specific requirements and constraints of each application scenario.

Differential Privacy has been integrated into Empirical Risk Minimization (ERM)

[1] census.gov/newsroom/press-releases/2021/2020-census-key-parameters.html

to balance the trade-offs between privacy and utility, with the optimal selection of DP parameters being studied [Li et al. 2023]. Additionally, Influence Functions (IF) have been employed to estimate the impact of DP mechanisms on model performance without the need for costly retraining [Carey et al. 2024]. These approaches address the challenge of parameter selection in DP and Table 1 compares these related works to our approach.

Table 1. Comparison between related works and this study

Reference	Independent variables	Values tested	Assessment
[Asquith et al. 2022]	Privacy budget	$\varepsilon \in \{0.25, 1.0, 8.0\}$	Accuracy
[Zhang et al. 2023]	Privacy budget, dataset size	$\varepsilon \in [0.1, 3]$ (13 values), $k \in [0.1, 1.0]$ (10 values)	Accuracy
[Li et al. 2023]	Privacy budget	$\varepsilon \in [0, 10]$	Utility in ERM
[Carey et al. 2024]	Privacy budget, group size	$\varepsilon \in [0.01, 10]$ (30 values), $k \in [0.01, 0.3]$ (10 values)	Utility via IF
Our work	Privacy budget, sensitivity	$\varepsilon \in [0.1, 12]$ (120 values), $\Delta f \in [0.1, 50.1]$ (51 values)	Distributional shift and accuracy

Among the studies compared in Table 1, ours is the only to use sensitivity (Δf) as an independent variable to assess DP. It also explores the largest set of values, considering all possible combinations of the independent variables. Ultimately, it is the only study to evaluate DP in the context of distributional shift. These distinctions emphasize the unique contributions and comprehensive scope of our work.

3. Materials and Methods

This study uses a dataset consisted of synthetically generated personal information simulating data leaks across 16 hypothetical adversarial disclosure scenarios [Sharma and Bantan 2025]. These scenarios cover a variety of sectors, including banking, e-commerce, food delivery, travel booking, healthcare, social networking, dating apps, and entertainment platforms. Each scenario incorporates relevant personally identifiable information to mimic realistic breach conditions. This approach allows the evaluation of differential privacy techniques without compromising the confidentiality of actual individuals.

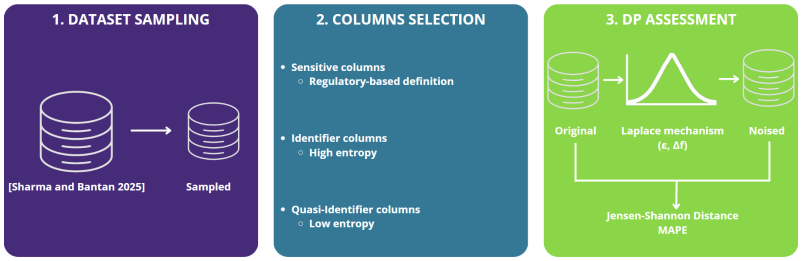


Figure 1. The study workflow

The methodology employed in this research is illustrated in Figure 1, and the corresponding steps are further elaborated in Sections 3.1 to 3.4. The analysis was conducted using Python version 3.11.11.

3.1. Dataset sampling

Due to the large size of the datasets, sampling is employed to reduce computational resource requirements while maintaining the representativeness of the entire population and ensuring statistical validity. In Equation 1, which is used to calculate the required sample size, Z represents the Z-score, corresponding to the desired confidence level; p denotes the estimated proportion; and e stands for the margin of error [Cochran 1977]. The variable n represents the sample size.

$$n = \frac{Z^2 \cdot p \cdot (1 - p)}{e^2} \tag{1}$$

A sample size of 185,000 rows is used across all adversarial disclosure scenarios, resulting in a margin of error of 0.3% with a 99% confidence level (corresponding to a Z-score of 2.576). This calculation assumes a proportion of 0.5, which represents maximum variability and provides the most conservative estimate. To ensure a random selection of samples, thereby meeting the requirement for unbiased sampling, we utilize the Unix command `shuf`.

3.2. Columns selection

For the experiment, a selection of columns to which differential privacy will be applied is made. To facilitate this process, the dataset columns are classified into three categories: sensitive attributes, identifiers and quasi-identifiers (QIs).

3.2.1. Sensitive attributes

The selection of sensitive columns is based on regulatory requirements. In the absence of a federal privacy law in the United States, we reference the California Consumer Privacy Act (CCPA) and the California Privacy Rights Act (CPRA), which amended the CCPA. These laws define sensitive attributes as any personal information that could expose a consumer's credentials, such as access to an account, financial account, debit card, or credit card number. Furthermore, the Payment Card Industry Data Security Standard (PCI-DSS) specifies that the Card Verification Value (CVV) should never be stored after a transaction is completed.

Based on these definitions, the following columns are considered sensitive attributes: username, e-mail address, password, credit card number, CVV and card expiry date. Since the protection of this information requires absolute confidentiality rather than statistical indistinguishability, these columns are excluded from the differential privacy assessment. The last four digits of the credit card number, however, are not considered sensitive, as PCI-DSS permits their storage as long as they are properly masked.

3.2.2. Identifier and quasi-identifier attributes

An identifier is an attribute that uniquely identifies an individual without the need for additional information, whereas a quasi-identifier can identify an individual when combined with other QIs, but not in isolation. All columns not classified as sensitive are categorized as either identifiers or quasi-identifiers. To categorize an attribute as an identifier or QI, we use entropy $H(X)$, as defined in Equation 2, where $P(x_i)$ represents the probability of the outcome x_i occurring [Shannon 1948].

$$H(X) = -\sum_{i=1}^{n} P(x_i) \ln P(x_i) \tag{2}$$

For identifiers, each value is unique and, thus, $P(x_i)$ follows a uniform distribution, with all values having a probability of $\frac{1}{185{,}000}$, based on the sample size. This results in an entropy of approximately 12.13 nats. Consequently, all columns with entropy greater than 12 nats are considered identifiers. Using this method, columns such as name, physical address, phone number and payment history are classified as identifiers. The columns that are neither considered sensitive nor identifiers are classified as quasi-identifiers. Identifiers are excluded from the evaluation, as any distortion applied to them would uniformly affect all values, maintaining their uniqueness. As a result, the differential privacy mechanism is applied exclusively to quasi-identifiers.

3.3. Adversarial disclosure scenario selection

Out of the 16 breach scenarios presented in [Sharma and Bantan 2025], we select the hotel booking breach (scenario 5) because it resulted in the largest number of quasi-identifiers according to the described selection methodology, thereby facilitating a more comprehensive assessment. This choice also aligns with previous research, which has indicated that the hotel industry often adopts inadequate and outdated security measures to protect its assets [Prabhu et al. 2023]. Additionally, the tourism sector faces numerous cyber threats [Ghaderi et al. 2024], further justifying the selection of this scenario.

Table 2. Columns categorized as QI and considered in the DP assessment [Sharma and Bantan 2025]

Column	Description
Device Information	The phone operating system (either Android or iOS) and its version, considering those released between 2016 and 2022
Travel Habits	Can be any combination of Road Trips, Train Travel, Cruises, Domestic Flights, International Flights
Payment Methods	Can be any combination of Cash, Online Wallet, Mobile Payment, Bank Transfer, Debit Card, Credit Card
City	Subject's address city (USA)
Card Last4digits	The last four digits of the credit card number
Zip Codes	Subject's address ZIP code (USA)

The selected adversarial disclosure scenario consists of 2.9 million rows (pre-sampling), with the columns classified as quasi-identifiers presented in Table 2. This

Table 3. Summary statistics of the QI columns

Column	Count	Unique Values	Most Frequent	Frequency
Device Information	185,000	16	Android, Android 12	11,773
Travel Habits	185,000	155	Train Travel	12,646
Payment Methods	185,000	3,905	Credit Card	7,590
City	185,000	17,610	Washington	1,444
Card Last4digits	185,000	10,000	**** **** **** 5519	36
Zip Codes	185,000	39,359	51342	16

table lists the key QI columns considered for the differential privacy assessment, along with brief descriptions of each attribute. The QI columns include indirectly identifying information that, when combined with other data, could potentially re-identify an individual, but not necessarily on its own. Table 3 provides a summary of the statistics for these QI columns. It includes the total count of entries for each column, the number of unique values, the most frequent value and its frequency of occurrence.

3.4. Differential privacy assessment

The `diffprivlib` library is employed to apply differential privacy to the dataset [Holohan et al. 2019]. It is chosen for its user-friendly Python API, comprehensive documentation and built-in warnings that assist in preventing inadvertent privacy violations [Ngong et al. 2024]. This library offers implementations of various differential privacy mechanisms, including Gaussian, Staircase and Bingham. For this study, we use the Laplace mechanism, which is widely adopted in privacy-preserving data analysis due to its simplicity and effectiveness [Li and Wang 2024].

3.4.1. The Laplace mechanism for differential privacy

The Laplace mechanism for differential privacy ($\mathcal{M}_L$), as proposed by [Dwork et al. 2016] and defined by Equation 3, ensures ε-differential privacy by adding calibrated noise to the output of a query. This noise is generated from a Laplace distribution, with its scale parameter being inversely proportional to the privacy budget ε.

$$\mathcal{M}_L(x, f, \epsilon) = f(x) + \text{Lap}\left(0, \frac{\Delta f}{\epsilon}\right) \tag{3}$$

Given an original query $f(x)$ with L_1-sensitivity Δf, the mechanism outputs $\mathcal{M}_L(x)$, where $\text{Lap}(0, b)$ denotes Laplace-distributed noise with mean 0 and scale $b = \Delta f/\epsilon$ determined by both the sensitivity Δf and the privacy budget ε. The larger the sensitivity Δf, the more noise is added to the original query's output. This leads to a higher level of privacy but also a greater loss of accuracy in the query's results. Conversely, a smaller sensitivity results in less noise, providing more accuracy but less privacy. Similarly, the privacy budget ε controls the trade-off: a smaller ε increases privacy by adding more noise, while a larger ε reduces privacy by adding less noise.

In this study, we examine how distributional shifts in query results are affected by changes in both sensitivity (Δf) and privacy budget (ε) for a count query $f(x)$. We

explore variations in privacy loss by adjusting $\varepsilon \in \{0.1, 0.2, 0.3, \ldots, 12.0\}$, and sensitivity over $\Delta f \in \{0.1, 1.1, 2.1, \ldots, 50.1\}$. Since negative counts are not meaningful in the context of this query, all outputs are clipped to a minimum value of 0 to ensure validity.

3.4.2. Comparison metrics

To compare the count distributions of each QI column before and after applying differential privacy, we use the Jensen-Shannon Distance metric [Flovik 2024], which is the square root of the Jensen-Shannon Divergence. The JSD is calculated from the Kullback-Leibler Divergence (KLD), which is defined by Equation 4 and also called relative entropy, quantifies the expected discrepancy in surprise when assuming Q instead of the actual distribution P [Kullback and Leibler 1951]. It is an unbounded (i.e. $D_{\mathrm{KL}}(P \parallel Q) \in [0, +\infty)$), asymmetric statistical distance (i.e. $D_{\mathrm{KL}}(P \parallel Q) \neq D_{\mathrm{KL}}(Q \parallel P)$). It is 0 if and only if P and Q are identical. Due to the fraction $\frac{P(x_i)}{Q(x_i)}$, $D_{\mathrm{KL}}(P \parallel Q)$ tends to infinity if there are zeroes in the Q distribution.

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{i=1}^{n} P(x_i) \log \left(\frac{P(x_i)}{Q(x_i)} \right) \tag{4}$$

The Jensen-Shannon Distance (Equation 5) quantifies the discrepancy between two distributions, P and Q, by computing the square root of the average Kullback-Leibler divergence between each distribution and their midpoint distribution $M = \frac{P+Q}{2}$ [Lin 2002]. Unlike KLD, JSD is a bounded (i.e., $\mathrm{JSD}(P \parallel Q) \in [0, \sqrt{\ln(2)}]$), symmetric measure. It attains 0 if and only if P and Q are identical, and reaches its maximum ($\sqrt{\ln(2)}$) when P and Q are disjoint. Since JSD relies on the mixture distribution M, it naturally avoids the infinite divergence problem of KL. Thus, no small constant is added to null values.

$$D_{\mathrm{JS}}(P \parallel Q) = \sqrt{\frac{1}{2} D_{\mathrm{KL}}\left(P \parallel M\right) + \frac{1}{2} D_{\mathrm{KL}}\left(Q \parallel M\right)} \tag{5}$$

In this work, we use JSD due to its bounded output, allowing a normalized comparison between different pairs of distributions, thus enabling the measurement of the distributional shift. To evaluate the accuracy degradation, we use the Mean Absolute Percentage Error (MAPE), which measures the error of perturbed data relative to the original values. For a dataset with n elements, where y_i represents the original value and $\hat{y}_i$ is its noised counterpart, MAPE is computed as Equation 6.

$$\mathrm{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \% \tag{6}$$

4. Results and Discussion

This section presents the results obtained from the experiments, with a focus on quantifying the distributional shifts introduced by the application of differential privacy. As

a basis for the subsequent analysis and discussion, the entropy values for each quasi-identifier column in the dataset is calculated. These entropy values provide information regarding the variability of each QI attribute, which influence the impact of noise addition under distinct differential privacy configurations.

4.1. Quasi-identifier columns entropy

The entropy for each selected quasi-identifier column was calculated and is presented in Table 4. Correlating this information with the summary statistics in Table 3, it is noted that columns with lower entropy tend to have fewer unique values. This relationship originates from the definition of entropy as a measure of unpredictability. When the number of unique values in a column is low, the diversity of possible outcomes is reduced, making the data more predictable and, consequently, lowering its entropy.

Table 4. The entropy for each QI column

Column	Entropy (nats)
Device Information	2.78
Travel Habits	4.32
Payment Methods	6.42
City	8.93
Card Last4digits	9.18
Zip Codes	10.48

Columns with lower entropy, such as Device Information, present a reduced risk of re-identification because many records share the same values. This commonality lowers the likelihood that an individual can be uniquely distinguished, potentially requiring less stringent privacy protection. However, low-entropy columns also present challenges for differential privacy. Due to the limited variability in such columns, the noise added may not be sufficient to obscure outliers effectively. As a result, rare or unique values may still stand out after the noise is applied, reducing the overall privacy protection in these cases.

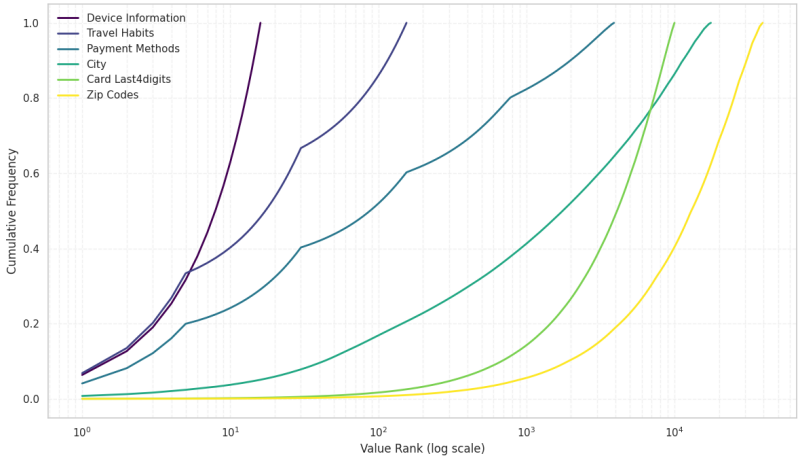


Figure 2. ECDF of value rank distribution of QI columns

Conversely, high-entropy columns, such as Zip Codes, have a greater potential for re-identification due to their higher uniqueness across records. However, because these

columns already exhibit a high degree of variability, even a small amount of noise can be effective in creating a distributional shift.

To assess the generalizability beyond the specific breach scenario analyzed, the Empirical Cumulative Distribution Functions (ECDFs) of the quasi-identifier attributes are estimated, as shown in Figure 2. This analysis reveals how representative these attributes are when compared to distributions typically found in real-world datasets. For instance, attributes such as Device Information and Travel Habits show highly concentrated distributions, consistent with patterns observed in many operational datasets, where a few dominant categories account for the majority of occurrences. Conversely, attributes like Zip Codes exhibit highly dispersed, long-tailed distributions. This is in accordance to Tabel 4 and suggests that, although our experiments were based on a specific scenario, the diversity of distributional patterns captured in our dataset makes it a reasonable representation for broader real-world applications.

4.2. Distributional shifts assessment

A distributional shift refers to a change in the statistical properties of the input data. In contexts such as machine learning, where differential privacy is often employed to protect the privacy of data subjects [Blanco-Justicia et al. 2022], such shifts can have significant consequences. Specifically, when there is a mismatch between the data distribution used during training and the one encountered in real-world deployment, model performance may degrade. This occurs because the model is optimized based on the training distribution and may not generalize well to data with altered characteristics, potentially leading to reduced accuracy and reliability during inference.

The distributional shifts as a function of the privacy budget and sensitivity, measured using the Jensen-Shannon Distance, are presented in Figure 3. As expected, the figure shows that greater shifts occur when more noise is added to the data. This happens when the privacy budget ε is lower and the sensitivity Δf is higher, both of which result in greater perturbation and thus increased divergence from the original distribution. This emphasizes the fundamental trade-off in differential privacy: adding more noise strengthens privacy guarantees but increases distributional shift, as indicated by the higher JSD values. Selecting appropriate values for ε and Δf involves carefully balancing these competing objectives. For example, smaller values of ε may be preferred in high-risk scenarios where stronger privacy protection is required, whereas larger ε can be tolerated for less sensitive attributes to preserve a better accuracy. Consequently, the choice of parameters must be guided by both the privacy requirements and the analytical goals of the specific application.

The variation patterns observed in Figure 3 can be directly linked to the entropy levels of each column. Columns with higher entropy, such as Payment Methods, City, Card Last4digits, and Zip Codes (Figures 3c-3f), exhibit higher Jensen-Shannon Distance values, even when minimal noise is introduced (i.e., at higher ε and lower Δf). This behavior reflects their greater variability, making their distributions more sensitive to even small perturbations introduced by differential privacy mechanisms. Because these attributes contain a larger number of distinct values and exhibit less uniform distributions, they are more sensitive to the injection of noise. As a result, the perturbations introduced by differential privacy mechanisms lead to more pronounced shifts in their probability

distributions. This increased sensitivity is reflected in higher Jensen-Shannon Distance values, even when the noise magnitude is relatively small.

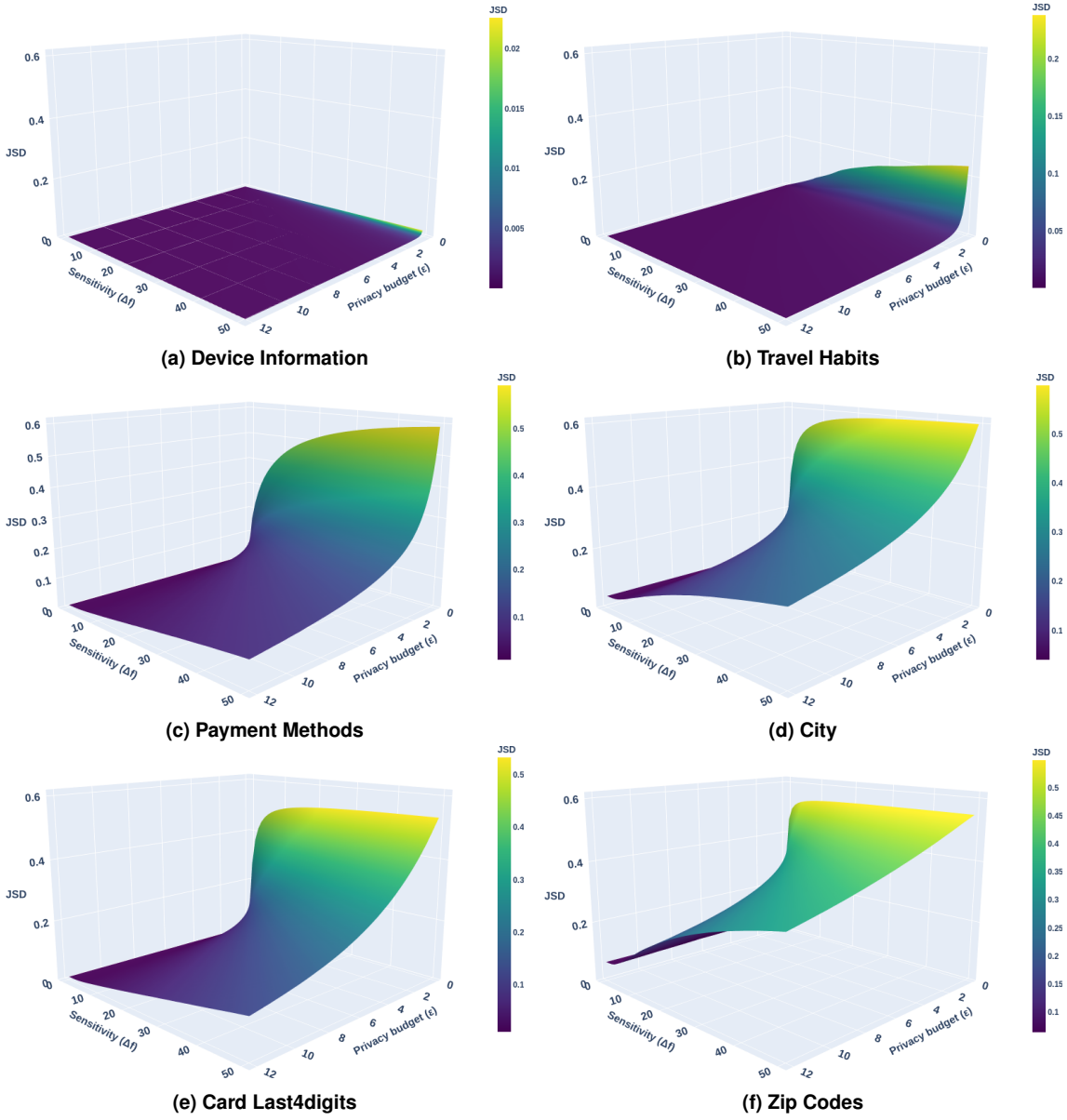


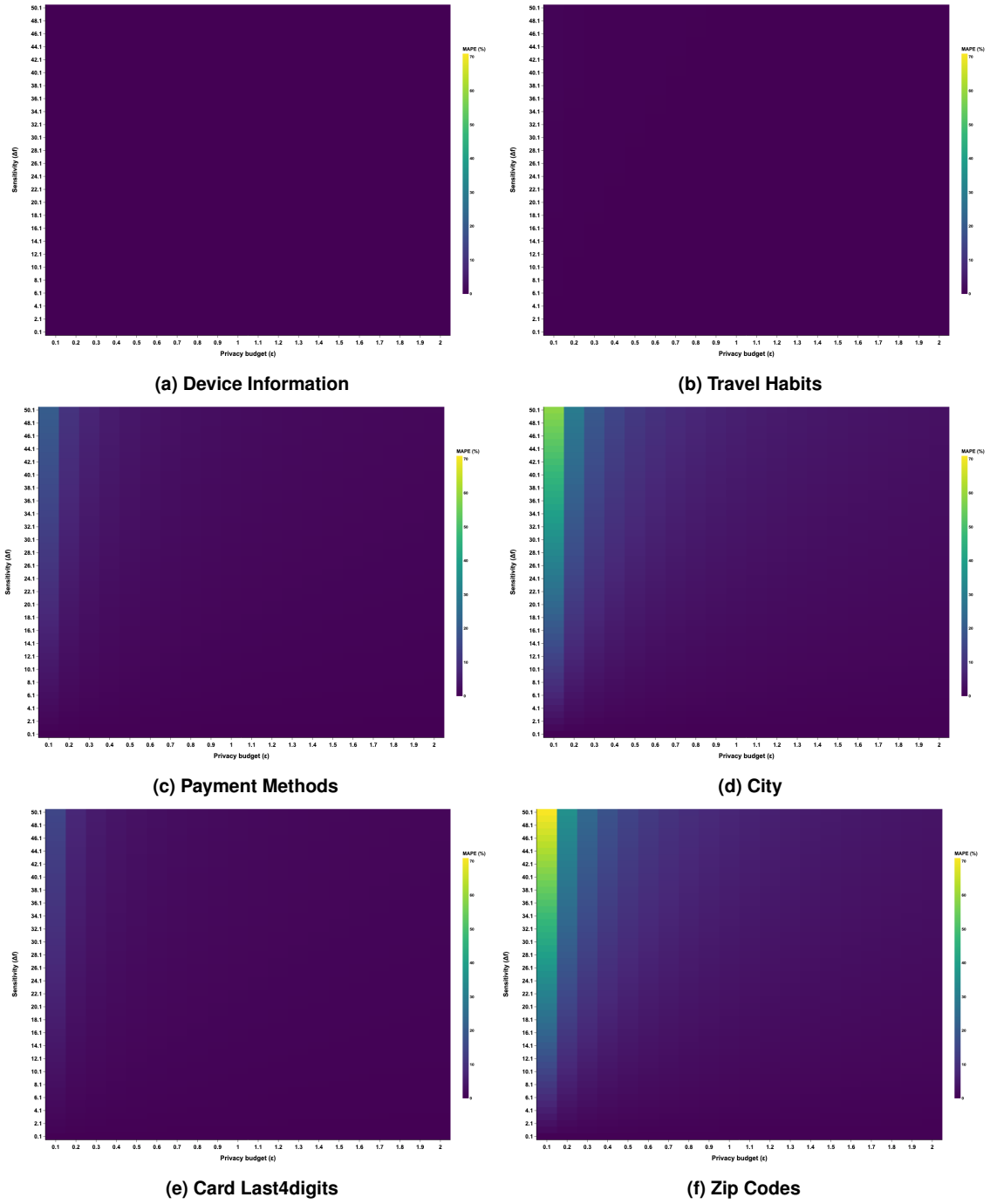
Figure 3. Jensen-Shannon Distances between the original and the noised datasets for different columns.

In contrast, columns such as Device Information and Travel Habits (Figures 3a-3b) consistently show low divergence values. These features exhibit lower entropy, as shown in Table 4, due to the smaller number of unique values, as evidenced in Table 3. This results in a minimal impact from noise injection, leading to smaller distributional shifts and consequently lower divergence, as reflected in the nearly flat surfaces in the corresponding plots. These columns, therefore, require significantly more noise (i.e., lower ε and/or higher Δf) to achieve the same level of privacy protection as higher-entropy columns. It can be observed that the magnitude of the divergence correlates with the entropy of the attribute, suggesting that features with greater diversity are more

susceptible to distortion when privacy-preserving mechanisms are applied.

4.3. Accuracy assessment

To assess the impact of noise on data utility, we measure accuracy using the Mean Absolute Percentage Error between the original and noised values. As expected, lower privacy budgets and higher sensitivity consistently lead to higher MAPE values, reflecting the trade-off between privacy preservation and the fidelity of the data for analysis.



(a) Device Information **(b) Travel Habits**

(c) Payment Methods **(d) City**

(e) Card Last4digits **(f) Zip Codes**

Figure 4. MAPE between the noised and original distributions, for $\varepsilon \leq 2$.

This accuracy degradation follows a predictable pattern, allowing the selection of

ε and Δf parameters that strike a balance between privacy requirements and the acceptable level of data accuracy for each specific case. Figure 4 presents the MAPE between the original and differentially private data for each quasi-identifier, for $\varepsilon \leq 2$ to provide a clearer visualization. Combined with Table 4, which reports the entropy values, a relationship between entropy and error under differential privacy is noticeable.

Low-entropy attributes, such as Device Information (2.78 nats), experience consistently lower MAPE across different privacy levels (Figure 4a). Medium-entropy columns, like Payment Methods (6.42 nats), exhibit moderate MAPE (Figure 4c), while high-entropy columns, such as Zip Codes (10.48 nats), show higher MAPE (Figure 4f), particularly at lower ε values. These attributes are more prone to a significant degradation in accuracy, requiring less noise injection to induce substantial errors. Therefore, it is observed that higher entropy and sensitivity result in greater distortion for the same privacy budget.

Across the tested configurations, it is observed that ε values in the range of $[1.5, 4.0]$ typically resulted in moderate noise injection, balancing privacy with reasonable levels of accuracy degradation (MAPE $< 15\%$) for most quasi-identifiers. Similarly, sensitivity values in the range $[5.0, 20.0]$ preserved accuracy without excessively compromising privacy for medium to high entropy attributes.

While these ranges are not definitive, they represent starting points for real-world deployment in contexts with similar data structures. Importantly, attributes with higher entropy indicate the need to tune DP parameters per attribute characteristics rather than using uniform values.

5. Conclusions and Future Works

This study assessed the privacy-utility trade-off in differentially private data publishing using synthetic data from a hypothetical hotel booking adversarial disclosure scenario. We applied the Laplace mechanism to quasi-identifiers and analyzed the effects of varying the privacy budget (ε) and sensitivity (Δf). Our findings demonstrate the significant impact these parameters have on distributional shifts and accuracy. High-entropy attributes were found to be more sensitive to noise, leading to greater distributional divergence and accuracy degradation, while low-entropy attributes exhibited greater resilience to perturbations. These results emphasize the importance of adapting the differential privacy parameterization to attribute entropy and the specific sensitivity requirements of the use case.

We conducted experiments across 6,120 different scenarios, evaluating all possible combinations of the differential privacy parameters. Each scenario was assessed using both Jensen-Shannon Distance and MAPE to quantify the distributional shifts and accuracy degradation. These results aid the informed selection of differential privacy parameters, helping to achieve a balance between privacy protection and data accuracy.

The findings of this study provide significant knowledge for practitioners to implement differential privacy in real-world applications. For instance, the entropy-based parameter tuning approach proposed in this work may be used to select ε and Δf values that match the data sensitivity profile. The alignment of privacy configurations with attribute-level entropy enables the balance between privacy and accuracy more effectively, contributing to compliant and responsible data handling practices.

For future work, it is proposed to extend this assessment to other differential privacy mechanisms, such as Gaussian or Exponential noise and evaluate their impact in comparison to the Laplace mechanism. Additionally, exploring the applicability of these techniques across different datasets would provide a broader understanding of their effectiveness. Another possibility for future research would involve investigating the combination of differential privacy with other anonymization techniques, like k-anonymity or t-closeness, to further enhance privacy while maintaining data accuracy. Future works could also explore the effects on distributional shift with another metrics, such as Mahalanobis distance. Additionally, exploring data analysis techniques such as t-Distributed Stochastic Neighbor Embedding (t-SNE) or Uniform Manifold Approximation and Projection (UMAP) visualizations could help reveal how the injected noise alters the structure of the dataset. Another direction involves training machine learning models, such as autoencoders, on the differentially private (noised) data and evaluating their performance on clean (non-noised) test sets.

References

Asquith, B., Hershbein, B., Kugler, T., Reed, S., Ruggles, S., Schroeder, J., Yesiltepe, S., and Van Riper, D. (2022). Assessing the Impact of Differential Privacy on Measures of Population and Racial Residential Segregation. *Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/1rsg867y.

Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., and Muralidhar, K. (2022). A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Computing Surveys*, 55(8):1–16.

Bohr, A. and Memarzadeh, K. (2020). Chapter 2 - the rise of artificial intelligence in healthcare applications. In Bohr, A. and Memarzadeh, K., editors, *Artificial Intelligence in Healthcare*, pages 25–60. Academic Press.

Cai, Y., Zhang, Y., Qu, J., and Li, W. (2022). Differential privacy preserving dynamic data release scheme based on jensen-shannon divergence. *China Communications*, 19(6):11–21.

Canedo, E. D., Bandeira, I. N., Calazans, A. T. S., Costa, P. H. T., Cançado, E. C. R., and Bonifácio, R. (2023). Privacy requirements elicitation: a systematic literature review and perception analysis of IT practitioners. *Requir. Eng.*, 28(2):177–194.

Carey, A. N., Van, M.-H., and Wu, X. (2024). Evaluating the impact of local differential privacy on utility loss via influence functions. In *2024 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2016). Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality*, 7(3):17–51.

Flovik, V. (2024). Quantifying distribution shifts and uncertainties for enhanced model robustness in machine learning applications. *arXiv preprint arXiv:2405.01978*.

Garcia, R. D. and Ueyama, J. (2024). Blockchain-based data governance for privacy-preserving in multi-stakeholder settings. In *Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg)*, pages 33–40. SBC.

Ghaderi, Z., Beal, L., and Houanti, L. (2024). Cybersecurity threats in tourism and hospitality: perspectives from tourists engaging with sharing economy services. *Current Issues in Tourism*, pages 1–16.

Holohan, N., Braghin, S., Mac Aonghusa, P., and Levacher, K. (2019). Diffprivlib: the IBM differential privacy library. *arXiv preprint arXiv:1907.02444*.

Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E. T., Simko, T., and Imai, K. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 us census. *Science advances*, 7(41):eabk3283.

Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86.

Lei, J., Wang, L., Pei, Q., Sun, W., Lin, X., and Liu, X. (2024). Privgrid: Privacy-preserving individual load forecasting service for smart grid. *IEEE Transactions on Information Forensics and Security*, 19:6856–6870.

Li, N. and Wang, T. (2024). Review of popular algorithms for differential privacy. In *Handbook of Sharing Confidential Data*, pages 39–51. Chapman and Hall/CRC.

Li, Y., Liu, Y., Li, B., Wang, W., and Liu, N. (2023). Towards practical differential privacy in data analysis: Understanding the effect of epsilon on utility in private erm. *Computers & Security*, 128:103147.

Lin, J. (2002). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151.

Malinin, A., Band, N., Ganshin, Alexander, Chesnokov, G., Gal, Y., Gales, M. J. F., Noskov, A., Ploskonosov, A., Prokhorenkova, L., Provilkov, I., Raina, V., Raina, V., Roginskiy, Denis, Shmatova, M., Tigas, P., and Yangel, B. (2022). Shifts: A dataset of real distributional shift across multiple large-scale tasks. *arXiv*.

Mir, D. J. (2012). Information-theoretic foundations of differential privacy. In *International symposium on foundations and practice of security*, pages 374–381. Springer.

Mueller, J. T. and Santos-Lozada, A. R. (2022). The 2020 us census differential privacy method introduces disproportionate discrepancies for rural and non-white populations. *Population Research and Policy Review*, 41(4):1417–1430.

Ngong, I. C., Stenger, B., Near, J. P., and Feng, Y. (2024). Evaluating the usability of differential privacy tools with data practitioners. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 21–40, Philadelphia, PA. USENIX Association.

Ouadrhiri, A. E. and Abdelhadi, A. (2022). Differential privacy for deep and federated learning: A survey. *IEEE Access*, 10:22359–22380.

Prabhu, B. A., Dani, R., and Bhatt, C. (2023). A study of the challenges faced by the hotel sector with regards to cyber security. In *Automation and computation*, pages 284–294. CRC Press.

Prokhorenkov, D. (2022). Anonymization level and compliance for differential privacy: A systematic literature review. In *2022 International Wireless Communications and Mobile Computing (IWCMC)*, pages 1119–1124.

Rigaki, M. and Garcia, S. (2023). A survey of privacy attacks in machine learning. *ACM Comput. Surv.*, 56(4).

Rodrigues, G. A. P., de Oliveira, M. N., Serrano, A. L. M., Rocha Filho, G. P., Vergara, G. F., Mosquéra, L. R., and Gonçalves, V. P. (2025). MELISSA: An LLM-powered smart home energy consumption monitoring framework. In *Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP)*, pages 11–20. SBC.

Sato, T. and Minamide, Y. (2025). Differential privacy. *Arch. Formal Proofs*, 2025.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Sharma, A. and Bantan, M. (2025). Simulating data breaches: Synthetic datasets for depicting personally identifiable information through scenario-based breaches. *Data in Brief*, 58:111207.

Tomás, J., Rasteiro, D., and Bernardino, J. (2022). Data anonymization: An experimental evaluation using open-source tools. *Future Internet*, 14(6).

Wairimu, S., Iwaya, L. H., Fritsch, L., and Lindskog, S. (2024). On the evaluation of privacy impact assessment and privacy risk assessment methodologies: A systematic literature review. *IEEE Access*, 12:19625–19650.

Wen, M., Xie, R., Lu, K., Wang, L., and Zhang, K. (2022). Feddetect: A novel privacy-preserving federated learning framework for energy theft detection in smart grid. *IEEE Internet of Things Journal*, 9(8):6069–6080.

Yao, A., Li, G., Li, X., Jiang, F., Xu, J., and Liu, X. (2023). Differential privacy in edge computing-based smart city applications:security issues, solutions and future directions. *Array*, 19:100293.

Zhang, S., Hagermalm, A., Slavnic, S., Schiller, E. M., and Almgren, M. (2023). Evaluation of open-source tools for differential privacy. *Sensors*, 23(14).