Beyond Data Collection: Safeguarding User Privacy in Social Robotics

DORAFSHANIAN, Mahboobeh, AITSAM, Muhammad, MEJRI, Mohamed and DI NUOVO, Alessandro <http://orcid.org/0000-0003-2677-2650>

Beyond Data Collection: Safeguarding User Privacy in Social Robotics

DORAFSHANIAN, Mahboobeh, AITSAM, Muhammad, MEJRI, Mohamed and DI NUOVO, Alessandro <http://orcid.org/0000-0003-2677-2650>

Beyond Data Collection: Safeguarding User Privacy in Social Robotics

Mahboobeh Dorafshanian
Department of Computer Science & Software Engineering
Laval University, QC, Canada
mador146@ulaval.ca

Muhammad Aitsam
Department of Computing
Sheffield Hallam University
Sheffield, UK
m.aitsam@shu.ac.uk

Mohamed Mejri
Department of Computer Science & Software Engineering
Laval University, QC, Canada
momej@ulaval.ca

Alessandro Di Nuovo
Department of Computing
Sheffield Hallam University
Sheffield, UK
a.dinuovo@shu.ac.uk

***Abstract*—In an era marked by the advanced capabilities of social robots in personal and public spaces, the issue of pervasive data collection by these entities becomes increasingly pertinent. Social robots, deployed by government entities, hospitals, and corporations, are at the forefront of gathering sensitive personal data, necessitating careful consideration of privacy concerns. The vast amounts of data collected by these robots, while beneficial for decision-making and fostering research, also pose significant privacy risks. In particular, the challenge intensifies when robots collect and potentially share data that includes sensitive personal information. This paper presents a user-friendly Differential Privacy (DP) library that addresses this challenge. The library incorporates a risk threshold and evaluates the potential impact of data disclosure to accurately quantify privacy levels. Designed for nontechnical users, it enables the secure release of statistical data without the risk of privacy breaches. With privacy breaches and re-identification becoming increasingly common, this library offers a robust solution for safeguarding individuals' privacy while facilitating the sharing of valuable insights.**

***Index Terms*—differential privacy, social robotics, sensitive data, data disclosure**

I. Introduction

In the evolving field of social robotics, where robots interact closely with individuals in settings ranging from homes to healthcare facilities, privacy concerns have become increasingly critical. The extensive personal data collected by these robots, ranging from behavioral patterns to personal preferences, intensifies the need for stringent data security and confidentiality measures. As these robots become more integrated into daily activities, both for personal use and in broader institutional settings, the imperative for robust privacy protection mechanisms grows. Historical instances of data breaches and the inadequacy of traditional privacy methods have highlighted the urgency for a paradigm shift in data privacy approaches, particularly in the context of social robotics, where the stakes of personal privacy are significantly heightened [1]. One pivotal milestone in this transformative journey was the introduction of Differential Privacy by Cynthia Dwork [2] and her colleagues. At its core, Differential Privacy is a revolutionary concept that re-frames the discourse on privacy in the digital era. It offers a principled and rigorous framework to balance the conflicting objectives of data analysis and individual privacy protection. Unlike traditional methods that often rely on anonymization or data aggregation, Differential Privacy introduces a mathematical approach that guarantees privacy, even in the presence of powerful adversaries. Differential Privacy is founded on a simple, yet powerful principle: the impact of any individual's data on the output of a computation should be minimal. To formalize this concept, Dwork proposed a mathematical definition of differential privacy as follows [3]:

Definition 1: A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ε,δ)-differential private if for all $\mathcal{S} \subseteq Range(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$Pr[\mathcal{M}(x) \in \mathcal{S}] \leq exp(\varepsilon) Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

In simple terms, a computation is differentially private if the probability of obtaining a particular result remains roughly the same, whether a specific individual's data is included or excluded. The key mechanism employed to achieve Differential Privacy involves the intentional introduction of controlled noise into the data or its analysis. This noise serves to mask the contribution of any single individual, ensuring that the presence or absence of their data does not unduly influence the overall result.

Figure 1 shows the Differential Privacy framework where the difference in the computation results of two neighbouring dataset could be at most of ε. This ε is also known as the privacy parameter. The privacy parameter (ε) quantifies the level of privacy protection, with smaller values of epsilon

This work is funded by Marie Sklodowska-Curie Action Horizon 2020 (Grant agreement No. 955778) for the project 'Personalized Robotics as Service Oriented Applications (PERSEO)'. This research is also supported by the Beneva Insurance and Natural Sciences and Engineering Research Council of Canada (NSERC).

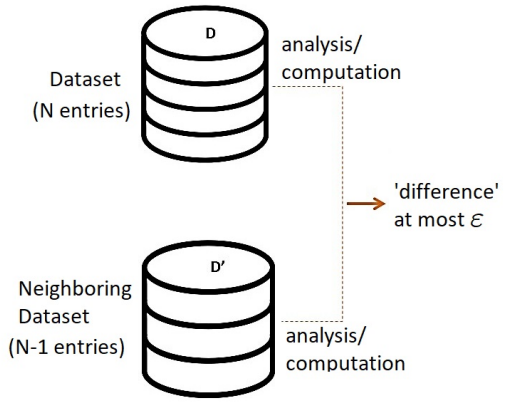

Fig. 1. Differential Privacy framework. Here we have two datasets, first is complete dataset and second is where information of one individual is removed or replaced. After computation, the maximum difference between the results is at most of ε.

indicating stronger privacy guarantees. This paper navigates the intricacies of Differential Privacy, delving into its fundamental definitions, underlying mechanisms, and implications for the broader landscape of data privacy. As the implementation of Differential Privacy poses challenges, especially for individuals without a technical background in this domain, understanding the paradigm's significance becomes crucial.

The goal of this paper is to present an algorithm designed to assist users in estimating the privacy parameter (ε). This parameter, fundamental to the inclusion of controlled noise in datasets, is crucial for those seeking to make the data collected by the robot, differentially private before releasing stats about it. Recognizing that many individuals lack the technical expertise required to apply Differential Privacy algorithms to their data, our algorithm seeks to simplify this process. By asking the user basic details, we aim to estimate the privacy parameter (ε), contributing to a more accessible and user-friendly approach to Differential Privacy.

The remainder of this paper is organized as follows. Section 2 addresses privacy breaches in the past, examining existing Differential Privacy algorithms and libraries. Section 3, details the methodology of our proposed algorithm. The experimental results are discussed in Section 4. Finally, the paper concludes in Section 5.

Our contributions to this paper are:

- Simplified Differential Privacy library that can easily be used by non-experts in this area.
- Estimation of the privacy parameter ε by taking into account the risk threshold (R_T) and the impact of data disclosure (I).

Through this exploration, our goal is to not only contribute to the ongoing dialogue on privacy in the field of social robotics but also to highlight the instrumental role that Differential Privacy plays in shaping a more secure and ethically grounded data ecosystem.

II. Background

The growing trend of data sharing has raised privacy concerns, as research suggests that combining seemingly innocuous information such as date of birth, gender, and zip code can uniquely identify a significant portion of the US population. In Canada, the risk of information disclosure has increased dramatically by 160% annually. The California Consumer Privacy Act (CCPA) [4] and the General Data Protection Regulation (GDPR) [5] have been introduced to provide rigorous privacy guarantees for users when analyzing and collecting data. Fraudsters stole 16.8 billion from US consumers in 2017 [6], posing a significant risk to users' personal information. The vulnerability of privacy preservation models to background knowledge attacks further complicates data protection. Several high-profile privacy failures have fuelled ongoing debates about these issues.

The AOL Privacy Debacle in 2006, the release of medical records by the Insurance Commission (GIS), and the Netflix Prize competition, all exposed the risk of privacy breaches despite claims of anonymization. These incidents demonstrated that even anonymized data can compromise privacy when linked with publicly available information. Researchers were able to re-identify many subscribers in the Netflix dataset by linking it with external databases such as The Internet Movie Database (IMDB). Other instances of re-identifying users from vehicle sensor data and extracting signals from logs also underscore the persistent risk of privacy breaches even with anonymized data.

These privacy failures highlight the urgent need for enhanced privacy techniques. Differential Privacy emerges as a potent solution, capable of neutralizing the aforementioned attacks and preventing such breaches by introducing controlled noise into datasets. By incorporating Differential Privacy, users participating in studies can be assured that their privacy will not be compromised, offering a crucial safeguard in the era of escalating privacy concerns.

A. Preliminaries

Differential Privacy aims to hide the participation of individuals. One of the basic approaches is the Laplace mechanism which adds the noise taken from the Laplace distribution [7].

Definition 2: Laplace Mechanism
Given any function $f : N^{|x|} \rightarrow R^k$, the Laplace mechanism is define as:

$$M_L(x, f(.), \epsilon) \;=\; f(x) + (Y_1, \ldots, Y_k)$$

where Y_i are i.i.d random variables drawn from Lap($\frac{\triangle f}{\epsilon}$).

By this definition the Laplace mechanism is (ε,δ)-differential privacy or ε-differentially private, where δ is always equal to 0. The Laplace mechanism is commonly used for numeric queries like count, sum, mean and variance.

B. Open-Source DP Libraries

Practical applications in the field of Differential Privacy have lagged behind theoretical advancements, with a notable scarcity of practical implementations. IBM-diffprivlib

TABLE I
COMPARISON OF DIFFERENTIAL PRIVACY LIBRARIES. [7], [8]

Parameters / Libraries	Owner	Usecase	Programming Languages	Type of Mechanisms	Type of Queries	ε Managment	Calculating Sensitivity
SmartNoise [9]	Microsoft	Data science and large systems	Python	Laplace, Gaussian, Exponential, Geometric	Count, Sum, Mean, Var, Histogram, Max, Min, Median	Yes	Yes
Google DP [10]	Google	Production ready applications	C++, Java, Go	Laplace, Gaussian, Exponential	Count, Sum,Mean, Var, Histogram, Max, Min, Median	Yes	Yes
Diffprivlib [11]	IBM	Data science	Python	Laplace, Gaussian, Exponential, Geometric, Binary, Vector	Count, Sum, Mean, Var, Histogram	Yes	Yes
Diffpriv [12]	B. Rubinstein et al.	Data science	R	Laplace, Gaussian, Exponential,	Any, provided the sensitivity sampler	N/A	Yes
Chorus [13]	J. P. Near et al.	large Scale systems	Scala	Laplace, Gaussian, Noisy Max, FLEX,SVT, Aggregate	Count, Sum, Mean, Histogram	Yes	Yes
GRAM-DP [14]	M.Aitsam	large Scale systems	Python	Laplace	Count, Sum, Mean, Variance	Yes	Yes

[11], OpenDP-Smartnoise [9], Openmined-PyDP [10], Diffpriv [12], Pinq, TensorFlow-Privacy, and Opacus-PyTorch, are few open-source frameworks that claim to validate global Differential Privacy for specific datasets. In one of comparative study G.Garrido [8] investigates seven open-source libraries.

Microsoft in collaboration with Harvard University (OpenDp initiative) provides an open-source library. SmartNoise [9] is a tool for accounting for the privacy budget, APIs for performing DP analysis, releasing statistics, and measuring the utility of the outputs. Although SmartNoise is designed for large-scale systems, it does not perform well for multiple datasets. Since one database must follow a special framework and others use, e.g., Sklearn, it fails machine learning implementations.

OpenMined provides the Google DP library [10], which is licensed under Apache-2.0 and supports Python and C++ libraries. Although experts can use the library directly, the Apache Beam tool provides a layer for non-expert users. Google DP offers advantages such as tracking privacy budgets, estimating analytical sensitivity bounds, and using a stochastic tester that does not break the DP guarantee. However, Google DP does not include machine learning algorithms.

The IBM general-purpose Diffprivlib library [11], implements in Python. It is based on the worst-case, so the development of the mechanisms requires theoretical sensitivity analysis. It uses formal mechanisms such as Exponential, Gaussian, or Laplace, and also includes the Vector [15], the Staircase [16], and Geometrics [17]. Its disadvantage is not having a floating point. On the other hand, running the Sklearn classifiers gives the ability to track the privacy budget in this library.

Diffpriv library [12] provides new approaches to execute user-defined functions with R language. Moreover, its sampler can determine the sensitivity of a user-defined function according to the bounded Differential privacy. Of this advantage, the sensitivity does not need to be calculated by non-expert users. However, this new library cannot determine the privacy budget of accountants, and computing the sensitivity would be expensive even for simple queries.

Chorus library in Scala language [13], uses the existing mechanisms with collaborating SQL databases. This library consists of three main components: accounting for the privacy budget, a query rewriting frame that modifies a query before execution, and the framework for query analysis which calculates the sensitivity of the query. The important note is that this library is a framework for research and is not a complete system for deployment. Moreover, Chorus has less built-in mechanisms in comparison to other Libraries.

However, to effectively utilize these libraries, users must possess a thorough understanding of Differential Privacy, including its parameters and the operational mechanisms of the chosen library. Once invoked, the library generates differentially private results based on the specified parameters. For newcomers, comprehending these factors may pose a significant challenge. To address the aforementioned problems, one of the open-source libraries offers a solution. GRAM-DP is a library that requires users to specify their desired privacy level. The framework then takes care of the rest, returning differentially private results. In this paper, we are introducing the updated version of GRAM-DP, known as GRAM-DP 2.0.

III. GRAM-DP 2.0

GRAM-DP 2.0 is written in Python 3, a popular language for data analysis. Utilizing the functionality of the well-known NumPy package, the library ensures instant recognition of functions, and default parameters make it accessible to all users. Released under the MIT Open Source license, GRAM-DP is free to use and modify, welcoming user contributions to enhance its features. The library focuses on providing basic queries for statistical analysis and essential building blocks for differential privacy and handling noise addition. Its primary goal is to make Differential Privacy easily understandable, especially for those new to the concept. In its second release, the library is emphasizing the accurate estimation of ε (privacy parameter) based on specific queries, data sensitivity, and user requirements on the risk and impact of data disclosure. In this release, the user has to provide desired privacy and have to answer some qualitative questions related to the data. Through these values, the framework will calculate the risk threshold

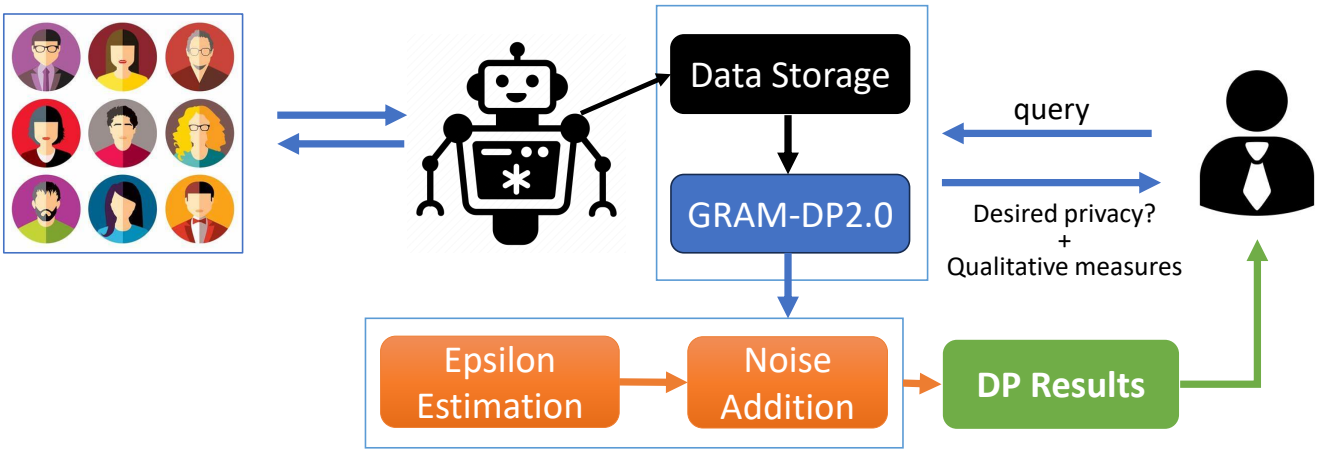

Fig. 2. Usage of GRAM-DP2.0 for social robotics. The social robot interacts with multiple users and saves their personal data. When a user queries about these data, GRAM-DP2.0 will make these data differentially private and return DP results.

TABLE II
QUERY AND EQUATION FOR SENSITIVITY [9]

Query	Sensitivity
Count	1
Sum	$M - m$
Mean	$(M - m)/n$
Variance	$((M - m)^2) * (n/(n^2 - 1))$

(R_T) and impact of data disclosure (I) which will eventually be used to estimate the ε. For sensitivity, we used the equations proved in the Harvard privacy tools project [9]. Here M and m are the maximum and minimum of data respectively, and n is the size of the data.

IV. Privacy Parameter (ε) Estimation

Epsilon (ε) or privacy budget, is the most important parameter in Differential privacy which controls the level of privacy. In most practical Differential privacy case studies, the appropriate value of ε is hardly calculated, so researchers prefer to consider fixed values for ε. Moreover, decision-makers often struggle to comprehend the significance of this key parameter. They typically base their decisions on the potential risks and consequences involved. They may have established risk thresholds that guide their decision-making process. The subsequent theorem establishes a connection between the privacy budget (ε), the risk threshold (R_T), and the impact of data disclosure (I).

A. Risk and Privacy Budget

Privacy budget (ε) is intuitively formulated, by using the confidence probability of the noise estimation.

Theorem 1: [18] If ξ is the number of values in the estimated distribution and the $\max(Lap(\frac{1}{\lambda})) \geq \frac{\xi-1}{2}$, then we can formulate ε as follow:

$$\max(Lap(\tfrac{1}{\lambda})) = -\frac{\varepsilon \times ln(2-2\gamma)}{\Delta q} \geq \frac{\xi-1}{2}$$

$$\Rightarrow \quad \frac{\varepsilon \times ln(2-2\gamma)}{\Delta q} \leq \frac{1-\xi}{2}$$

$$\Rightarrow \quad \varepsilon \leq \frac{\Delta q(1-\xi)}{2 \times ln(2-2\gamma)}$$

Here, ε is estimated according to its relationship with the risk of data disclosure (RoD). The following Theorem gives an upper bound for the privacy budget ε based on I and R_T.

Theorem 2: [7] Let q be a query and I be the impact of its privacy disclosure. Let R_T be a risk threshold (the maximum risk that the company can tolerate). The privacy budget ε with Laplace noise needs to be equal to or less than

$$u \times (1 - \frac{I}{R_T})$$

where $u = \frac{\Delta q}{2 \times ln(2-2\gamma)}$.

This Theorem is for single-dimensional data. Now we generalize the theorem to n queries.

Theorem 3: [7] Let $q_1, \ldots, q_n$ be n queries and $I_1, \ldots, I_n$ be the impacts of their privacy disclosures, respectively. Let R_T be a risk threshold (the maximum risk that the company can tolerate). The global privacy budget ε with Laplace noise is equal to or less than

$$U - \frac{\sum_{i=1}^{n} u_i \times I_i}{R_T}$$

where $U = \sum_{i=1}^{n} u_i$ and $u_i = \frac{\Delta q_i}{2 \times ln(2-2\gamma)}$.

The proofs of Theorems 1 and 2 are described in [7]. Recognizing that many individuals lack the technical expertise required to apply differential privacy algorithms to their data. We use this simplified notion for ε along with the GRAM-DP 2.0 library to facilitate understanding Differential Privacy.

V. Methodology

To assess our proposed solution, we substitute the value of ε by R_t and I. Then we use GRAM-DP 2.0 for three datasets, namely Adult Dataset [19], Diabetes Health Indicator Dataset [20], and Life Style Dataset [21] to get the differential private results. We chose the 'Age' column for our experiments as it is one of the most common data collected by robots. Another reason behind choosing this column is that it satisfies most of the queries (except *sum* query).

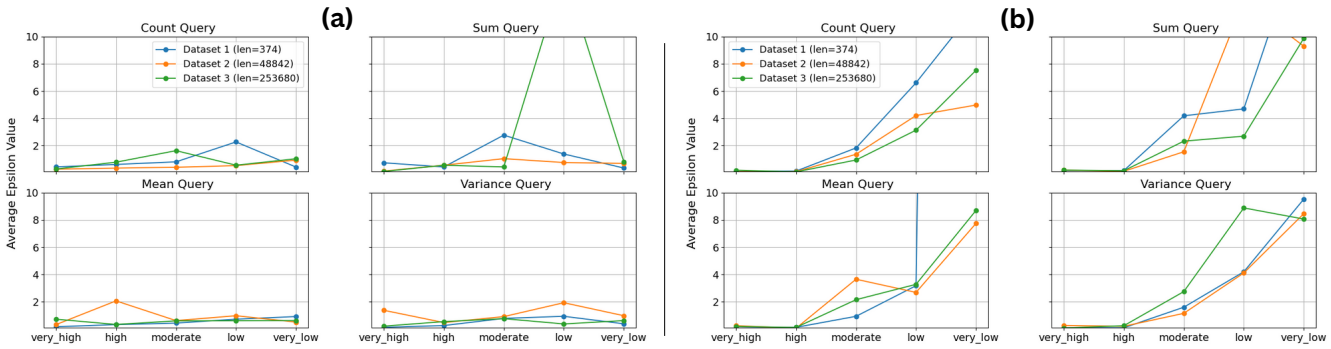

Fig. 3. Average estimated epsilon for each query.

Figure 2 shows the overall process. Here social robot interacts with multiple users and during interaction it also saves personal information to provide personalized responses in the next interaction. All of these data are stored in the robot's data storage unit. When the user (not malicious and wants to release data to the public) queries the robot for some information regarding the users, the GRAM-DP 2.0 will ask about its desired privacy and qualitative measures. From this information, GRAM-DP 2.0 will estimate the value for ε which will then be used to add noise to the query result and make it differentially private. The user will get this DP result for the query.

From the equation derived in section IV, to estimate ε value, we need sensitivity, Laplacian noise, Risk Threshold (R_T) and Impact of Data Disclosure (I).

1) Sensitivity: Table II provides the equations for the calculation of sensitivity, depending on the query.

2) Laplacian Noise: To calculate the Laplacian curve we need the initial value of epsilon. Instead of asking this value from the user, which could be difficult for him/her, we estimated this value from the user's answer to the question of desired privacy. If the user desires to have high privacy then the estimated value of the initial epsilon will be lower and vice versa. Note that this is an initial value of epsilon ε which is needed to calculate Laplace noise in the final ε equation.

3) Risk Threshold (R_T): Risk Threshold depends on the desired privacy level. The user (e.g. company's manager) has to tell the desired privacy level (very high, high, moderate, low, very low).

4) Impact of Data Disclosure (I): Understanding the impact of one result on another is hard to estimate. It depends on various factors like background knowledge, source, size, relevance, and sensitivity of the dataset. In this work, we estimated this value by taking into account two factors. 1) Qualitative measures where the user will be asked a series of general questions related to the dataset. 2) Utilizing the Identity Ecosystem Report [22] as a reference to know the importance of each attribute in the dataset. The average score of these factors is our total score for I. Refer to our GitHub (https://github.com/aitsam12/GRAM-DP2.0.git) for more details about these factors.

VI. Results and Discussion

We used three open-source datasets: Dataset 1 represents Life Style Dataset, Dataset 2 is the Adult Dataset, and Dataset 3 is the Diabetes Health Indicator Dataset. These datasets have varying sizes, making them suitable for evaluating GRAM-DP 2.0's performance on both small and large datasets.

Our main goal of this study is to estimate the privacy parameter ε. In the first experiment (Figure 3), for each dataset, we show the average ε of 100 runs for different privacy levels where the qualitative measures are maximum (a) and minimum (b). The Maximum level of qualitative measures indicates that the data is highly sensitive, while the lowest level of qualitative measures suggests that the data is not particularly sensitive. Figure 3(a) shows that when the data is sensitive, the value of epsilon is mostly low for all the queries which means that more noise is added to the true results. In Figure 3(b) where data is considered as less sensitive, less noise is added to the true results. This shows that GRAM-DP 2.0 complies with the basic definition of differential privacy.

Our proposed solution has been demonstrated to be consistent with the differential privacy definition. The next step is to determine the error in the data. In Figure 4, for every dataset, we took the average of 100 error values for each privacy level. Figure 4(a) shows the results for very sensitive data. Here for the count and sum query, the error is high till moderate privacy level. In the mean and variance query, we can see that the size of the dataset came into play. For small datasets, the error is higher as compared to large datasets. Similarly, in Figure 4(b) for less sensitive data, this trend follows for mean and variance queries. However, for count query error rate is almost similar irrespective of the size of the dataset. This is because of the nature of this query and its sensitivity value.

VII. Conclusion

Addressing the critical need for data privacy in the field of social robotics, we introduce GRAM-DP 2.0, a user-friendly Differential Privacy (DP) library. The proposed solution presents a dual advantage: it robustly safeguards individual privacy and simultaneously facilitates the sharing of

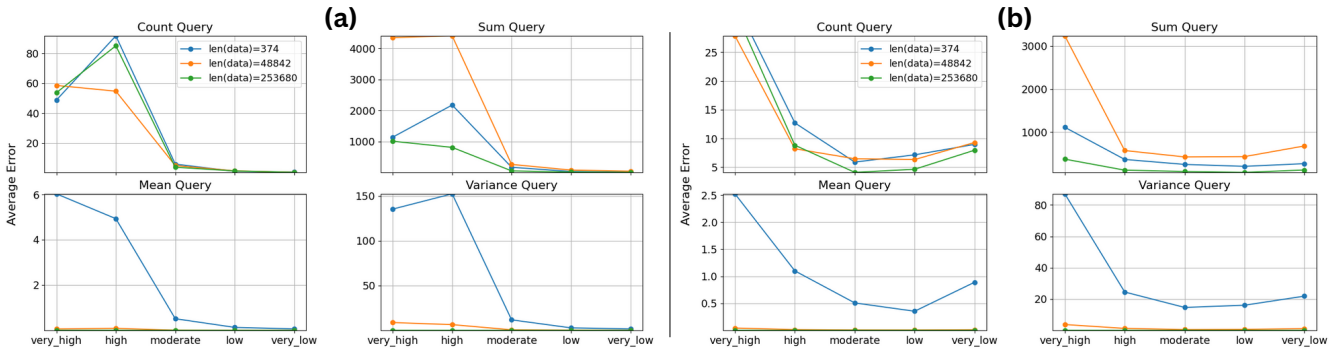

Fig. 4. Average Error for each query.

insightful data. The results show that GRAM-DP 2.0 adheres to the principles of global differential privacy, making it a reliable tool in environments where sensitive data is prevalent. Designed for ease of use by nontechnical individuals, GRAM-DP 2.0 ensures an accurate estimate of the privacy parameter (ε) by taking minimum input from the user. In essence, GRAM-DP 2.0 is a pivotal step toward balancing the benefits of social robotics with the imperatives of data privacy in the modern era.

ACKNOWLEDGMENT

This work is funded by Marie Sklodowska-Curie Action Horizon 2020 (Grant agreement No. 955778) for the project 'Personalized Robotics as Service Oriented Applications (PERSEO)'. This research also is supported by the Beneva Insurance and Natural Sciences and Engineering Research Council of Canada (NSERC). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] J. Marchang and A. D. Nuovo, "Assistive multimodal robotic system (amrsys): Security and privacy issues, challenges, and possible solutions," *Applied Sciences 2022, Vol. 12, Page 2174*, vol. 12, p. 2174, 2 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/4/2174/htm https://www.mdpi.com/2076-3417/12/4/2174

[2] C. Dwork and J. Lei, "Differential privacy and robust statistics," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, pp. 371–380.

[3] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy." *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.

[4] E. Goldman, "An introduction to the california consumer privacy act (ccpa)," *Santa Clara Univ. Legal Studies Research Paper*, 2020.

[5] G. D. P. Regulation, "General data protection regulation (gdpr)," *Intersoft Consulting, Accessed in October*, vol. 24, no. 1, 2018.

[6] A. Pascual, K. Marchini, and S. Miller, "identity fraud: Fraud enters a new era of complexity," *Javelin Strategy & Research. Report. Retrieved September*, vol. 30, p. 2019, 2018.

[7] M. Dorafshanian and M. Mejri, "Differential privacy: Toward a better tuning of the privacy budget (ε) based on risk." in *ICISSP*, 2023, pp. 783–792.

[8] G. M. Garrido, J. Near, A. Muhammad, W. He, R. Matzutt, and F. Matthes, "Do i get the privacy i need? benchmarking utility in differential privacy libraries," *arXiv preprint arXiv:2109.10789*, 2021.

[9] M. Gaboardi, M. Hay, and S. Vadhan, "A programming framework for opendp," *Manuscript, May*, 2020.

[10] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, 2014, pp. 1054–1067.

[11] N. Holohan, S. Braghin, P. Mac Aonghusa, and K. Levacher, "Diffprivlib: the ibm differential privacy library," *arXiv preprint arXiv:1907.02444*, 2019.

[12] B. I. Rubinstein and A. Francesco, "diffpriv: An r package for easy differential privacy," *Journal of Machine Learning Research*, vol. 18, pp. 1–5, 2017.

[13] N. Johnson, J. P. Near, J. M. Hellerstein, and D. Song, "Chorus: a programming framework for building scalable differential privacy mechanisms," in *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2020, pp. 535–551.

[14] M. Aitsam, "Differential privacy made easy," in *2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*. IEEE, 2022, pp. 1–7.

[15] R. Bassily, A. Smith, and A. Thakurta, "Private empirical risk minimization: Efficient algorithms and tight error bounds," in *2014 IEEE 55th annual symposium on foundations of computer science*. IEEE, 2014, pp. 464–473.

[16] Q. Geng and P. Viswanath, "The optimal mechanism in differential privacy," in *2014 IEEE International Symposium on information theory*. IEEE, 2014, pp. 2371–2375.

[17] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," in *Proceedings of the forty-first annual ACM symposium on Theory of computing*, 2009, pp. 351–360.

[18] Y.-T. Tsou, H.-L. Chen, and Y.-H. Chang, "Rod: Evaluating the risk of data disclosure using noise estimation for differential privacy," *IEEE Transactions on Big Data*, 2019.

[19] B. Becker and R. Kohavi, "Adult," UCI Machine Learning Repository, 1996, DOI: https://doi.org/10.24432/C5XW20.

[20] CDC, "Diabetes Health Indicator Dataset, https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators, 2014.

[21] J. M. Dzierzewski, S. M. Sabet, S. M. Ghose, E. Perez, P. Soto, S. G. Ravyts, and N. D. Dautovich, "Lifestyle factors and sleep health across the lifespan," *International Journal of Environmental Research and Public Health*, vol. 18, p. 6626, 6 2021. [Online]. Available: /pmc/articles/PMC8296445/ /pmc/articles/PMC8296445/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8296445/

[22] UCI repository, https://archive.ics.uci.edu/ml/datasets/adult, 1996.