A Seminar Report

Differential Privacy in Areas of Artificial Intelligence

Submitted in partial fulfillment of the requirements for the award of the degree+

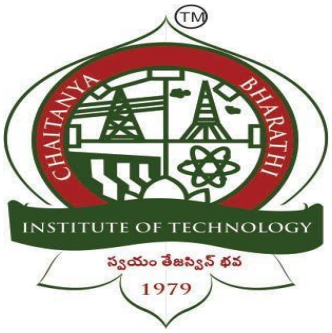
BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

by

SAMREEN SULTHANA (160119733316)



Department of Computer Science and Engineering

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)
(Affiliated to Osmania University; Accredited by NBA(AICTE) and NAAC(UGC), ISO Certified 9001:2015)
GANDIPET, HYDERABAD – 500 075
Website: www.cbit.ac.in
[2022-2023]

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)
Kokapet(Village), Gandipet, Hyderabad, Telangana-500075. www.cbit.ac.in

COMMITTED TO RESEARCH, INNOVATION AND EDUCATION 44 years

CERTIFICATE

Certified that seminar work entitled "**Differential Privacy in Areas of Artificial Intelligence"** is a bonafide work carried out in the eighth semester by "**Samreen Sulthana (160119733316)"** in partial fulfillment for the award of Bachelor of Engineering in Computer Science and Engineering from Chaitanya Bharathi Institute of Technology(A), Gandipet during the academic year 2022-2023.

SIGNATURE
Dr. Sugandha Singh

SIGNATURE
K. Kiran Prakash

SIGNATURE
Mr. A. Mohan

SIGNATURE
Mr. K. Karthik

INDEX PAGE

Topic Page **Page No**

TABLE PAGE

Table Name **Page No**

Image Index

Image Name **Page No**

ABSTRACT

Artificial Intelligence (AI) has seen tremendous progress in recent years, developing advanced algorithms and deep learning models. However, as AI continues to evolve, several challenges have emerged, including concerns about privacy violations, security issues, and model fairness. Differential privacy is a mathematical model that has emerged as an effective solution to these challenges, making it a valuable tool in AI.

Differential privacy offers several properties that can help preserve privacy and security in AI models, such as adding random noise to data sets and limiting the amount of information that can be extracted from them. It can also be used to improve the stability of learning, build fair models, and impose composition in selected areas of AI.

Regular machine learning, distributed machine learning, deep learning, and multi-agent systems are some of the areas where differential privacy can be applied to improve AI performance. In regular machine learning, differential privacy can help prevent overfitting and improve generalization. In distributed machine learning, it can ensure that sensitive data remains private while still allowing models to be trained. In deep learning, it can help address issues related to model transparency and fairness. Finally, multi-agent systems can help ensure that all agents have equal access to information.

Overall, differential privacy has emerged as a powerful tool for solving some of the most significant challenges facing the field of AI today. As AI continues to evolve, the importance of this tool will only continue to grow, helping to ensure that the benefits of AI can be realized while protecting individual privacy and security.

1. Introduction

Artificial Intelligence (AI) has become a popular area of research in recent years, with applications ranging from distributed control systems to mobile computing. However, as AI systems become more reliant on data, several new problems have emerged, including privacy violations, security issues, model instability, and fairness concerns. One tool that has gained attention in the AI community for addressing these issues is differential privacy. Differential privacy is a privacy preservation model that introduces calibrated randomization to the aggregate output, providing a level of privacy protection while maintaining data utility.

Differential privacy is a powerful privacy-preserving technique that has gained significant attention in the field of artificial intelligence. It has been widely applied in many areas, including deep learning, machine learning, and multiagent systems. The primary goal of differential privacy is to ensure that the sensitive information contained in the data used for training or inference is not disclosed to unauthorized parties. This technique adds a controlled amount of noise to the data to make it difficult for an attacker to learn anything about the individuals who contributed their data. In this way, differential privacy provides a way to balance the need for accurate results with the need for privacy protection. In this context, this paper discusses the application of differential privacy in deep learning, machine learning, and multiagent systems and highlights its benefits, limitations, and potential future research directions.

1.1. Areas of Artificial Intelligence

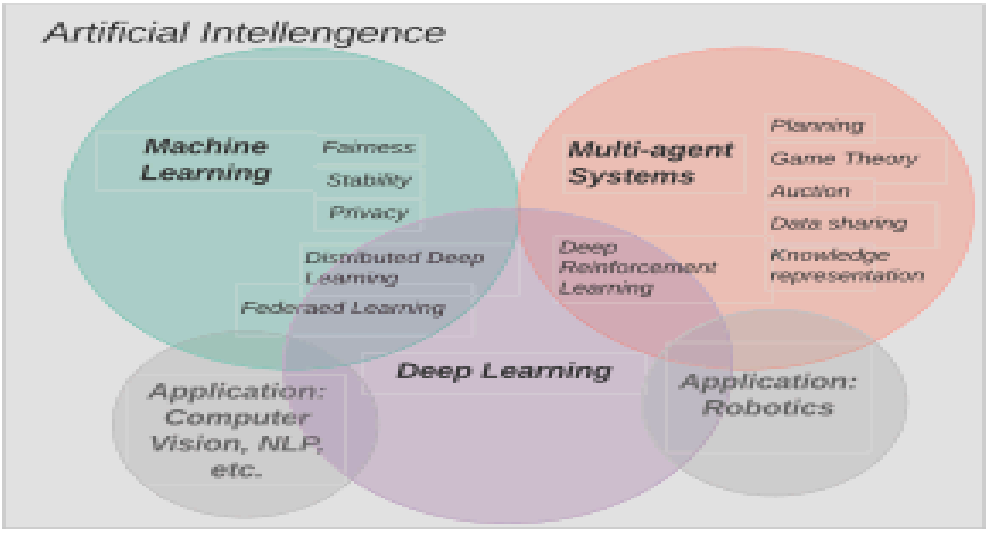


Fig-1: Areas of Artificial Intelligence

Deep Learning (DL), Machine Learning (ML), and Multiagent systems are three key areas of Artificial Intelligence (AI) that have gained significant attention in recent years.

DL is a subset of ML that uses neural networks to learn representations of data. It has been widely applied to tasks such as image and speech recognition, natural language processing, and even game playing. Some of the popular deep learning architectures include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs).

ML is a broader term that encompasses a wide range of algorithms and techniques that enable machines to learn from data and make predictions or decisions without being explicitly programmed. It has been applied to a wide range of problems such as fraud detection, recommender systems, and predictive maintenance.

Multiagent systems refer to a group of agents that interact with each other and with their environment to achieve a common goal. Each agent may have different capabilities, knowledge, and objectives, and they need to coordinate their actions to achieve the overall objective. Multiagent systems have been applied to a wide range of applications such as traffic management, supply chain management, and disaster response.

1.2. Differential Privacy in Areas of Artificial Intelligence

Differential privacy is a privacy-preserving mechanism increasingly used in various areas of artificial intelligence, including deep learning (DL), machine learning (ML), and multi-agent systems. The core idea of differential privacy is to ensure that the data of individual users remain private while allowing the analysis of the aggregate data.

In deep learning, differential privacy can be used to ensure the privacy of the training data while still allowing the model to learn from it. This is important because deep learning models are often trained on large datasets that may contain sensitive information, such as personal or medical data. By adding noise to the training data or the gradients used to update the model, differential privacy can prevent the disclosure of individual information while maintaining the model's accuracy as shown in Figure 2.

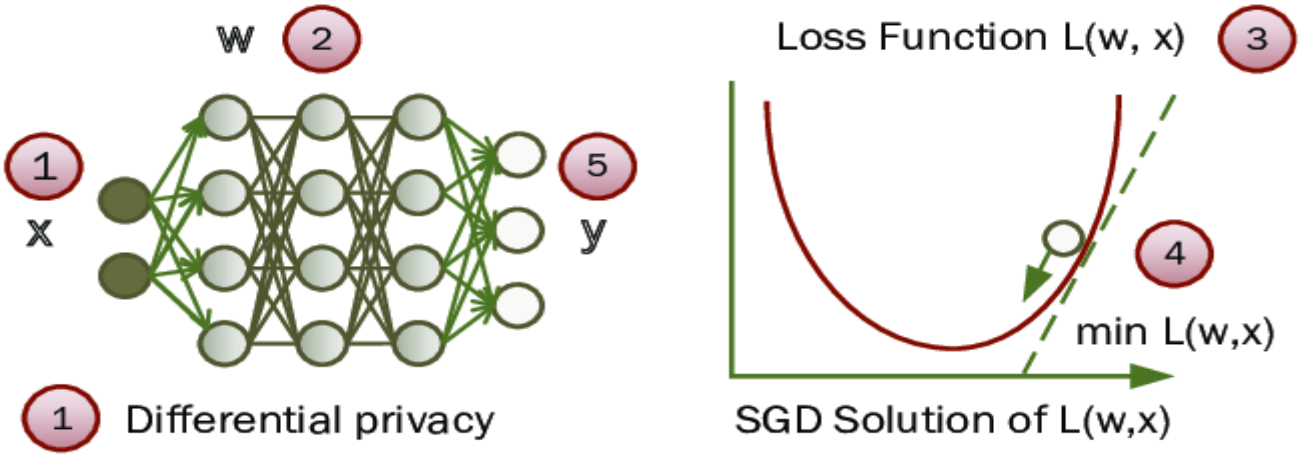


Fig-2: Differential Privacy for Deep Learning

In machine learning, differential privacy can be used to protect the privacy of individuals whose data is used to train the models. This is important because machine learning algorithms are often trained on data that contains sensitive information, such as health records or financial data. Differential privacy can be used to add noise to the data or the model's parameters to protect individual privacy while still allowing the model to learn from the data as shown in Figure-2.

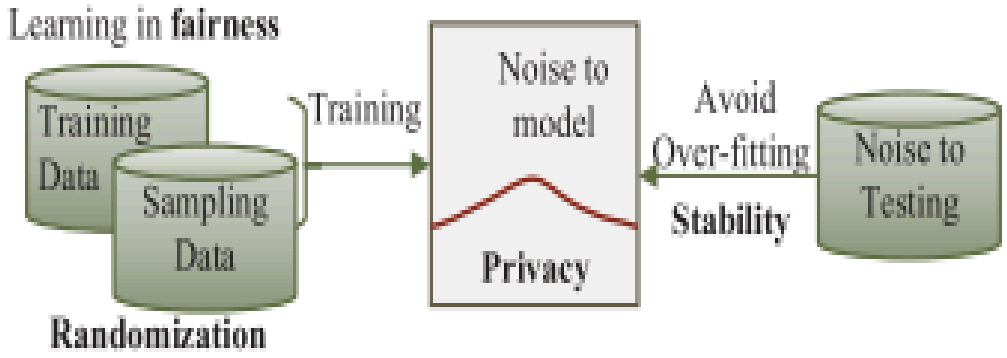


Figure-3: Learning Model

In multi-agent systems, differential privacy can be used to protect the privacy of individual agents while still allowing the system to learn from the aggregate data. This is important because multi-agent systems often involve multiple agents with different objectives and preferences, and protecting individual privacy can help ensure cooperation and fairness in the system.

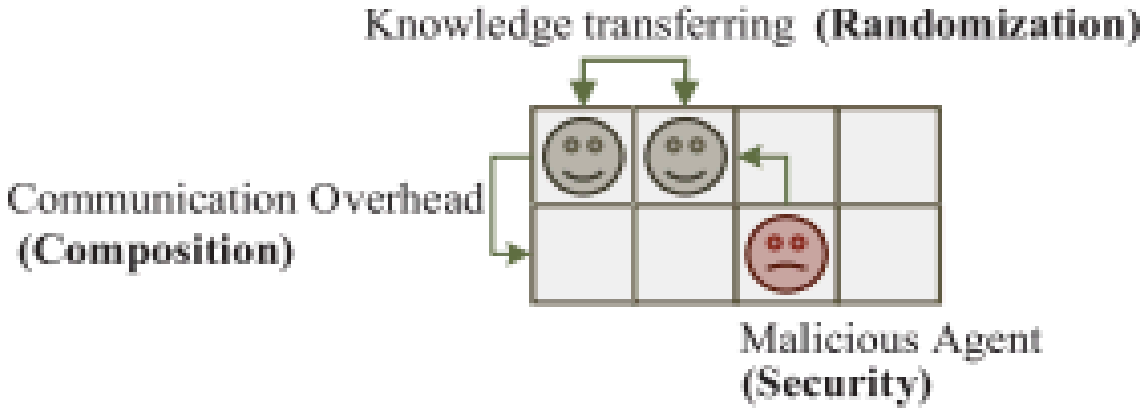


Figure-4: Multiagent systems

Overall, differential privacy is a powerful tool for ensuring privacy in various areas of artificial intelligence. Its use is likely to increase as more data is collected and used to train models and make decisions.

2. Literature Survey

2.1.More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence Tianqing Zhu, Dayong Ye, Wei Wang, Wanlei Zhou, and Philip S. Yu 2022.

The paper "More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence" by Tianqing Zhu et al. explores the potential applications of differential privacy in solving various problems in the field of artificial intelligence (AI). While AI has made great advancements in recent years, it has also encountered issues such as privacy violations, security concerns, and model fairness. Differential privacy is a mathematical model that can help address these issues and is thus a valuable tool in AI. However, no study has documented which differential privacy mechanisms have been used to overcome these problems. This paper shows that differential privacy can be used for more than just privacy preservation, and can also improve security, stabilize learning, build fair models, and impose composition in selected areas of AI such as regular machine learning, distributed machine learning, deep learning, and multi-agent systems. Overall, the article presents a new perspective on the potential for improving AI performance with differential privacy techniques.

This paper investigated the use of differential privacy in selected areas of AI, highlighting the critical issues facing AI and the basic concepts of differential privacy. The paper discussed how differential privacy can be applied to solve problems related to privacy violations, security issues, and model fairness. It also discussed the strengths and limitations of current studies in each area and pointed out potential research areas where the benefits of differential privacy remain untapped. The paper focused on machine learning, deep learning, and multi-agent learning, but the authors acknowledged that differential privacy has been used in other areas such as natural language processing, computer vision, and robotics.

2.2.C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in Proc. 3rd Conf. Theory Cryptogr. 2006.

The paper is discussing on builds on previous work on privacy-preserving statistical databases and proposes a method to protect privacy while allowing queries on general functions, not just noisy sums. The authors suggest adding calibrated noise to the database's true answer, which is the result of applying a query function to the database. The amount of noise added is based on the sensitivity of the function, which is the maximum amount the output can change due to the addition or removal of a single data point. By adding noise based

on sensitivity, the authors show that privacy can be preserved with less noise than previously thought, making the approach more practical.

To characterize privacy, the authors use the concept of indistinguishability of transcripts. Essentially, they show that two different queries should produce indistinguishable responses from the database, even if they differ only by a single data point. This ensures that an adversary cannot learn sensitive information about any individual from the responses.

The paper also highlights the advantages of interactive sanitization mechanisms over non-interactive ones. Interactive mechanisms allow for a dialogue between the user and the database to refine the query, leading to better privacy guarantees. Overall, the paper makes a significant contribution to the field of privacy-preserving statistical databases by proposing a more practical and efficient approach to protect privacy while allowing queries on general functions.

2.3.F. McSherry and K. Talwar, "Mechanism design via differential privacy," in Proc. 48th Annu. IEEE Symp. Found. Comput. Sci., 2007, pp. 94–103.

The paper explores the use of differential privacy in mechanism design, a field concerned with designing systems where participants have private information and incentives to act in their interest. The authors propose using differential privacy as a tool for mechanism design, allowing for the design of mechanisms that preserve privacy while ensuring desirable properties such as incentive compatibility and budget balance.

The paper presents several mechanisms designed using differential privacy, including a mechanism for selecting the most popular items among a group of participants and a mechanism for distributing a limited resource among participants. The authors analyze the privacy guarantees and incentive properties of each mechanism and show that differential privacy can be a useful tool in mechanism design.

2.4. F. McSherry, "Privacy integrated queries: An extensible platform for privacy-preserving data analysis," Commun. ACM, vol. 53, no. 9, pp. 89–97, 2010.

The paper proposes a new approach to privacy-preserving data analysis called Privacy Integrated Queries (PINQ). PINQ is an extensible platform that allows for different query types to be performed on data while preserving the privacy of the individuals whose data is being used.

The paper outlines the design and implementation of PINQ and shows how it can be used to perform various types of queries on different datasets while ensuring privacy. The platform incorporates differential privacy and allows for the addition of new query types through a modular architecture.

The paper also discusses the limitations and challenges of PINQ and presents potential areas for future research. Overall, the paper highlights the importance of privacy in data analysis and presents a promising new approach to preserving privacy while allowing for useful insights to be gained from data.

3. Methodology

3.1. Properties of Calibrated Randomization

Calibrated randomization benefits some AI algorithms. What follows is a summary of several properties derived from randomization.

1. Preserving privacy: This is the original purpose of differential privacy. By hiding an individual in the aggregate information, differential privacy can preserve the privacy of participants in a dataset.

2. Stability: Differential privacy mechanisms ensure that the probability of any outcome from a learning algorithm is unchanged by modifying any individual record in the training data. This property establishes connections between a learning algorithm and its ability to be generalized.

3. Security: Security relates to malicious participants in a system. Differential privacy mechanisms can reduce

the impact of malicious participants in AI tasks. This property can guarantee security in AI systems. Fairness. In machine learning, a given algorithm is said to be fair, or to have fairness, if its results are independent of sensitive attributes, like race and gender. Differential privacy can help to maintain fairness in a learning model by re-sampling the training data from the universe.

4. Composition: Differential privacy mechanisms can guarantee that any step that satisfies differential privacy can construct a new algorithm that also satisfies differential privacy. This property is referred to as composition and is controlled by the privacy budget.

Table-1: Properties of Differential Privacy in Artificial Intelligence

Selected AI areas		Privacy	Stability	Fairness	Security	Composition	Utility
Machine learning	Private learning	Yes				Yes	Decrease
	Stability in learning		Yes				Increase
	Fairness in learning			Yes			Increase
Deep learning	Deep Learning	Yes					Decrease
	Distributed deep learning	Yes				Yes	Decrease
	Federated learning	Yes		Yes		Yes	Decrease or Increase
Multi-agent system	Reinforcement learning	Yes			Yes	Yes	Increase
	Auction	Yes				Yes	Decrease
	Game theory					Yes	Decrease

In AI, the composition can be used to control the number of steps, communication loads, etc. Table 1 shows the properties that have been explored to date for each of our three disciplines. In machine learning, differential privacy has been applied to private learning, stability, and fairness. In deep learning, privacy is a major concern, but distributed deep learning and federated learning have also been investigated. In

multi-agent systems, differential privacy has been used to guarantee privacy, provide security, and ensure composition. The

utility shows the ultimate performance of the technology after adding differential privacy. Normally, privacy preservation comes with a utility cost. However, if differential privacy can contribute to stability or security, the utility may increase, such as in federated learning or fairness.

3.2. Mechanism

Definition 1:

d-Differential Privacy:

A randomized algorithm M gives -differential privacy for any pair of neighboring datasets D and D0, and, for every set of outcomes V, if M satisfies where Ω denotes the output range of the algorithm M.

$$\mathbf{Pr[M(D)\in\Omega]\leq exp(\in)\,Pr[M(D')\in\Omega]+\delta}$$

Definition 2:

Sensitivity:

For a query f : D →R, the sensitivity of f is defined as

$$\Delta f=\max_{D,\,D'}\|f(D)-f(D')\|1.$$

Two prevalent randomization mechanisms, Laplace and exponential, are used to satisfy the definition of differential privacy, but there are others, such as the Gaussian mechanism. Each is explained next.

3.2.1. Randomization: Laplace Mechanism

The Laplace mechanism is applied to numeric outputs. The mechanism adds independent noise to the original answer, as shown in Definition 3.

Definition 3:

Laplace mechanism: For a function f : D → R over a dataset D, the mechanism *M* in Eq. (3) provides ∈-differential privacy

$$M(D)=f(D)+Lap(\Delta f/\in)$$

3.2.2. Gaussian Mechanism

Compared to a Laplace mechanism, a Gaussian mechanism adds noise that is sampled from a zero-mean isotropic Gaussian distribution. The noise Z is sampled ~$N(0,\sigma^2)$ to the $L2$ sensitivity as follows:

$$\Delta f=\max_{D,D}\|f(D)-f(D')\|2$$

Definition 4:

Gaussian mechanism: For a function f : D → R over a dataset D, the mechanism *M* in Eq. (4) provides ∈ ; d-differential privacy

$$M(D)=f(D)+\sim N(0,\sigma^2),$$
$$\sigma=\Delta f\ \sqrt{2\log(1.25/\delta)}/\in$$

.

3.2.3. Exponential Mechanism

Exponential mechanisms are used to randomize the results for non-numeric queries. They are paired with a score function q(D, *Φ*) that evaluates the quality of an output f. Defining a score function is application-dependent, so different applications lead to various score functions.

Definition 5 (Exponential mechanism): Let q(D, *Φ*) be a score function of dataset D that measures the quality of output $\phi \in$ ϕ, Δq represents the sensitivity of *Φ*. The exponential mechanism *M* satisfies ∈-differential privacy if

M(D)=(return $\Phi \propto \exp(\in q(D,\phi)/2\Delta q)$).

3.2.4. Composition

Two privacy budget composition theorems are widely used in the design of differential privacy mechanisms: sequential composition [14] and parallel composition.

Theorem 1. Parallel Composition: Suppose we have a set of privacy steps *M={M1,.....Mm}*, if each *Mi* provides an ∈I privacy guarantee on a disjointed subset of the entire dataset, the parallel of *M* will provide max{∈1,. . . ,∈}-differential privacy.

Parallel composition corresponds to cases where each *Mi* is applied to disjointed subsets of the dataset. The ultimate privacy guarantee only depends on the largest privacy budget.

Theorem 2. Sequential Composition: Suppose a set of privacy steps *M={M1,.....Mm}* are sequentially performed on a dataset, and each *Mi* provides a privacy guarantee, *M* will provide (m.∈)-differential privacy.

Sequential composition offers a privacy guarantee for a sequence of differentially private computations. When a series of randomized mechanisms are performed sequentially on a dataset, the privacy budgets are added up for each step.

4. Applications

Differential Privacy is a technique used in data analysis and statistics to protect the privacy of individuals while still allowing for useful insights to be gleaned from large datasets. Some common applications of differential privacy include:

1. Statistical analysis: Differential privacy can be used to analyze large datasets while protecting the privacy of individuals. This is particularly useful in fields like healthcare and finance, where sensitive personal information is often involved.

2. Personalized recommendations: Differential privacy can be used to generate personalized recommendations for users while still protecting their privacy. For example, online retailers can use differential privacy to analyze user behavior and provide recommendations based on that behavior without compromising the privacy of the user.

3. Social network analysis: Differential privacy can be used to analyze social networks and identify patterns and trends in social behavior without revealing the identities of individual users.

4. Smart cities: Differential privacy can be used in the development of smart cities to protect the privacy of individuals while still allowing for data-driven decision-making. For example, traffic data can be analyzed to optimize traffic flow without revealing the movements of individual drivers.

5. Data sharing: Differential privacy can be used to enable secure data sharing between organizations without compromising the privacy of individuals. This is particularly useful in fields like healthcare and finance, where organizations need to share sensitive data to make informed decisions.

5. Conclusion

The key idea of differential privacy is to introduce calibrated randomization to the aggregate output. To show that applying differential privacy mechanisms to test data in machine learning could prevent the overfitting of learning algorithms, it launched a new direction beyond simple privacy preservation to one that solves emerging problems in AI.

The areas of AI where the benefits of differential privacy remain untapped. The areas of focus will be machine learning, deep learning, multi-agent learning, and areas of research in AI such as natural language processing, computer vision, robotics, etc.

6. Reference

[1].T. Zhu, D. Ye, W. Wang, W. Zhou and P. S. Yu, "More Than Privacy: Applying Differential Privacy in Key Areas of Artificial Intelligence," in *IEEE Transactions on Knowledge and Data Engineering,* vol. 34, no. 6, pp. 2824-2843, 1 June 2022, doi: 10.1109/TKDE.2020.3014246.

[2].Dwork, C., McSherry, F., Nissim, K., Smith, A. (2006). Calibrating Noise to Sensitivity in Private Data Analysis. In: Halevi, S., Rabin, T. (eds) Theory of Cryptography. TCC 2006. Lecture Notes in Computer Science, vol 3876. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11681878_14

[3].F. McSherry and K. Talwar, “Mechanism design via differential privacy,” in Proc. 48th Annu. IEEE Symp. Found. Comput. Sci., 2007, pp. 94–103.

[4].F. McSherry, “Privacy integrated queries: An extensible platform for privacy-preserving data analysis,” Commun. ACM, vol. 53, no. 9, pp. 89–97, 2010.