


Balancing Privacy and Performance: A Review of Differential Privacy in Deep and Federated Learning

Sardar Irfanullah Amanullah^{1*} , Sune von Solms² 

¹ Ph.D candidate, Electrical & Electronic Engineering Science, Faculty of Engineering & the Built Environment, University of Johannesburg, South Africa.

² Professor, Electrical & Electronic Engineering Science, Faculty of Engineering & the Built Environment, University of Johannesburg, South Africa.

*Corresponding Author: irsardar@gmail.com

ARTICLE INFO

ABSTRACT

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

The growing dependence on machine learning and deep learning models that are trained on sensitive user data has brought about important privacy issues that require strong solutions to prevent the disclosure of personal data. Differential Privacy (DP) has become a crucial technique for anonymizing individuals to prevent adversaries from identifying them from datasets. However, in distributed learning scenarios such as federated learning, the need for private and secure computation is even more critical. However, while DP approaches are rigorous from the privacy standpoint, their realization in practical systems is not always straightforward, especially for large and complex models such as deep neural networks. Such issues often result in a trade-off between the accuracy of the model and the computational cost, which limits the usability of DP in practical settings. This review aims to survey the usage of DP in deep learning and federated learning frameworks for the purpose of protecting personal data by adding noise to data models. The study reviews the core of DP, the probability distributions (Gaussian and Laplace) and how they are used in practice to bridge the gap between the theoretical and the real world. The comparison also brings out the major trends in privacy, accuracy, and robustness and reveals some major shortcomings in the ability to maintain model performance while achieving a high level of privacy. Furthermore, it discusses the problem of extending the DP concept to work with large datasets and how to control the level of noise such that it does not compromise the predictive capability of the model. This paper gives a detailed survey of the use of DP in improving the privacy of deep learning and the areas that need to be addressed for better implementation in challenging learning environments.

Keywords: Differential Privacy, Deep Learning, Federated Learning, Privacy Protection, Model Accuracy.

1. INTRODUCTION

Machine learning (ML), especially deep learning (DL), often uses large amounts of personal data for purposes like finance and healthcare. However, many of these systems need access to sensitive information, raising privacy concerns. In traditional ML systems, data is usually stored in central databases, which makes it vulnerable to leaks and unauthorized access. Distributed systems like federated learning reduce some privacy risks by allowing data to stay with users, but they still combine updates from multiple users, which can expose private information [1].

Protecting data privacy is now a key focus for researchers and professionals, both during the training of ML models and when making predictions. Differential privacy is an effective solution to these issues. It works by ensuring that adding or removing a single person's data from a dataset has little to no impact on the model's results [2]. This prevents attackers from figuring out specific details about individuals, even if they have access to the model's outputs or internal parameters.

DP is becoming a widely used method for privacy protection in both distributed systems like federated learning and traditional centralized systems. Its simple yet powerful guarantee makes it an essential tool for privacy-preserving machine learning.

Importance of Differential Privacy

Differential privacy is a mathematical method designed to ensure privacy by limiting how much an individual's data can influence the outcome of a model. It works by adding carefully calibrated noise to the model's computations, making it impossible to identify specific data points in the results. This ensures that attackers cannot extract information about individual users or reverse-engineer the data, as no single data point has a significant impact on the model's predictions [3].

Beyond protecting individual privacy, DP plays a critical role in fields like finance, healthcare and autonomous vehicles, where secure data sharing is essential. By protecting privacy, DP encourages data owners to share their information, leading to more accurate models and better predictions. Additionally, DP enables models to be trained directly on user devices without transferring sensitive data to a central server. This decentralized approach, common in federated learning, adds another layer of protection, making it harder for malicious actors to access private data [4].

As machine learning continues to expand across industries, the demand for privacy-preserving methods like DP has grown significantly. Regulations such as the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) push organizations to adopt privacy-conscious approaches, helping build trust in machine learning systems [5]. DP has been extensively researched and applied in various areas, including medical data analysis and recommendation systems, proving its effectiveness in both theory and practice [6].

Differential privacy offers robust privacy guarantees, striking an essential balance between protecting user data and ensuring the effectiveness of machine learning models. With ongoing research focused on improving its scalability, efficiency, and applicability across various tasks, DP is poised to be a fundamental tool in advancing privacy in artificial intelligence and machine learning. The following sections of this article provides a comprehensive exploration of DP, examining its key concepts and diverse applications in deep learning. Section 2 offers an overview of DP, highlighting its foundational principles and critical role in safeguarding privacy during data analysis. Section 3 explores DP's application in deep learning, emphasizing its importance in maintaining privacy during the training of complex models. Section 4 shifts focus to DP in federated learning, demonstrating its capacity to secure data privacy in decentralized settings. Section 5 examines key probability distributions used in DP, including Gaussian, Laplace, and Poisson distributions, which are vital for introducing noise and preserving privacy. In Section 6, we compare various DP variants, discussing their strengths, limitations, and optimal use cases. Section 7 delves into the theoretical and practical gaps within DP, addressing challenges such as the balance between privacy and utility and the need for enhanced methodologies. Lastly, Section 8 provides an outlook on the future of DP research, aiming to refine its effectiveness and expand its applications in privacy-preserving machine learning models.

2. UNDERSTANDING DIFFERENTIAL PRIVACY

Definition and Mathematical Foundations

DP is a mathematical framework designed to guarantee that a person's privacy is protected whenever their data is included in a dataset used for statistical analysis or machine learning operations. DP is essentially built on the concept that it assures, to a great extent, that the inclusion or deletion of a single data point inside a dataset does not materially affect the conclusion of a computation or analysis. We use a parameter called "epsilon" (ϵ) to measure the degree of privacy protection; this value controls the additional noise level to the results. Lowering the value of ϵ indicates that the privacy guarantees are more robust; nevertheless, it usually results in less accurate results [7]. The algorithm introduces a small amount of randomness, commonly known as noise, to ensure privacy protection while still allowing for effective analysis of the entire dataset.

It means that when using a random method M , the chance of getting a result in a set S should not change much if we use two datasets D and D' that are almost the same, except for one entry.

$$Pr[M(D) \in S] \leq e^\epsilon \cdot Pr[M(D') \in S]$$

where, ϵ represents the privacy budget, which controls the balance between privacy and utility. Smaller values of ϵ increase the noise, enhancing privacy protection. However, larger values of ϵ reduce the noise level but also increase the risk of exposing certain data points [8].

Mechanisms for Achieving DP

Several distinct techniques have been developed to achieve DP; each of them is ideal for a particular kind of analysis or machine learning model. The Laplace mechanism is among the most regularly used ones among them. It is a mechanism adding noise from the Laplace distribution to the query results so satisfying the DP criteria. The Laplace mechanism performs wonderfully for basic requirements such computation of average or sum. More advanced machine learning models demand additional procedures such as the Gaussian or exponential approaches. Each of these systems has specific advantages and applications [9].

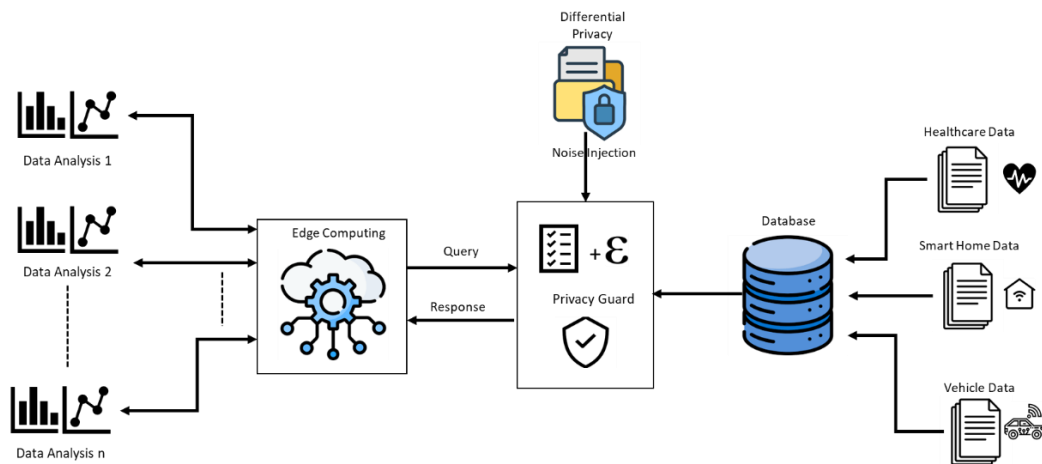


Figure 1: Differential Privacy Process in Mobile Edge Computing [10]

Moreover, Differentially Private Stochastic Gradient Descent (DP-SGD), routinely applied in the training stage of machine learning models under DP. DP-SGD is a modification on the respected stochastic gradient descent (SGD) technique. It introduces noise to the gradients all along the optimisation process. This ensures that the model does not leak important data when it changes. This approach is highly effective in training deep neural networks while preserving differential privacy. [11].

Noise Mechanisms (Gaussian, Laplace)

The Laplace Mechanism adds noise taken from the Laplace distribution, defined as:

$$Lap(\lambda) = \frac{1}{2\lambda} e^{-|x|/\lambda}$$

where λ is the scale parameter, determined by the privacy budget and sensitivity Δf of the function:

$$\lambda = \frac{\Delta f}{\epsilon}$$

This mechanism is particularly effective in scenarios requiring strict -Differential Privacy. The Laplace noise is symmetric and centered around zero, making it ideal for applications involving simple counting queries and discrete datasets [12], [13].

The Gaussian Mechanism adds noise from a Gaussian (normal) distribution:

$$N(0, \sigma^2)$$

Where σ is the standard deviation, chosen based on the privacy parameter (ϵ, δ) -DP and function sensitivity Δf :

$$\sigma = \frac{\Delta f \sqrt{2 \ln(1.25) / \delta}}{\epsilon}$$

Unlike the Laplace Mechanism, the Gaussian Mechanism satisfies (ϵ, δ) -Differential Privacy, making it suitable for applications requiring relaxed privacy constraints with better utility. This approach is widely used in machine learning and large-scale data analytics where stricter DP guarantees may reduce accuracy. Table 1 summarizes the core differences between the Laplace and Gaussian mechanisms in terms of their privacy guarantees, noise distribution, and the best use cases for each.

Table 1: Differences between the Laplace and Gaussian mechanisms

Mechanism	Privacy Guarantee	Noise Distribution	Best Use Cases
Laplace	ϵ -DP	Laplace (heavier tails)	Counting queries, histogram releases
Gaussian	(ϵ, δ) -DP	Normal (lighter tails)	Machine learning, large-scale data analysis

Laplace noise is more effective for small-scale discrete queries, whereas Gaussian noise provides better utility for complex computations in high-dimensional datasets.

The choice between the Laplace and Gaussian mechanisms depends on the privacy requirements, query sensitivity, and desired balance between privacy and accuracy. The Laplace Mechanism is suitable for strict DP requirements, while the Gaussian Mechanism provides flexibility in high-dimensional analyses [12], [13].

3. DP IN DEEP LEARNING

Application of DP in Neural Networks

DP applied in DL has proved about sensitive data to be a helpful technique for protecting individuals' privacy during the training deep learning process as seen in table 2. Deep learning models such neural networks depend on a lot of data to attain strong performance, which creates major privacy concerns. DP approaches these models using largely the injection of noise during the training phase. This is done to guarantee that the output of the model is not much affected by individual data points, hence restrict dissemination of private data.

Table 2: Application of DP in Neural Networks

Application	Description	Key Techniques/ Mechanisms	Challenges	Impact on Model Performance
Training Deep Neural Networks (DNNs)	Protecting individual data points while training deep models, such as CNNs, RNNs, and MLPs.	DP-SGD (Stochastic Gradient Descent), Gradient Noise Addition, Batch Noise Addition	High computational cost, loss of model accuracy due to noise, complex hyperparameter tuning	Moderate decrease in accuracy due to noise addition
Image Classification with DP	Ensuring privacy in image classification tasks with deep	Noise injection into gradients during backpropagation (e.g., Gaussian noise), clipping gradients	Risk of overfitting due to noisy gradients,	Small drop in accuracy compared to

	learning models like CNNs.		sensitivity of images to noise	non-privatized models
Privacy-Preserving Transfer Learning	Applying DP to transfer learning models, ensuring privacy during pre-trained model fine-tuning.	Differentially private fine-tuning, DP-SGD applied to transfer learning stages	Balancing between fine-tuning accuracy and privacy requirements	Lower accuracy in fine-tuning stages, but maintains privacy guarantee
Federated Learning with DP	Use of DP to protect user data in federated learning settings where data is decentralized across devices.	DP-SGD combined with federated learning frameworks, noise injection in local gradient computation	Challenges with heterogeneous data, communication overhead, ensuring privacy across multiple devices	Higher privacy, but potentially slower convergence and model performance
Adversarial Defense in Neural Networks	Using DP to mitigate adversarial attacks while training neural networks to be resistant to such threats.	DP combined with adversarial training, adding noise to the loss function and gradients to obfuscate adversarial input	Balancing robustness to adversarial attacks and maintaining privacy without compromising efficiency	Improved resistance to adversarial attacks but at the cost of performance
Privacy in Medical Data Analysis	Protecting sensitive medical data while training neural networks on large healthcare datasets.	DP-based data augmentation, applying noise to model parameters during training on medical images or health records	Ensuring clinical accuracy while maintaining strict privacy constraints	Lower accuracy, but ensures privacy of sensitive medical information
Natural Language Processing (NLP)	Applying DP to protect user privacy in NLP tasks such as sentiment analysis, machine translation, etc.	DP-SGD for word embeddings, noise addition during gradient descent on text data	High computational load, trade-off between preserving linguistic features and privacy	Reduced performance on NLP tasks with significant noise filtering

DP-SGD is among the most often utilised approaches for DP application in artificial intelligence. Adding noise into the gradient updates at each training iteration is the research of the ways this approach changes the traditional SGD

process. the research calibrates the noise considering the gradients of the sensitivity of the model and the intended degree of privacy (ϵ). Moreover, the use of noise guarantees that the gradients will not disclose too much information about any the research data point, therefore maintaining the personal privacy of every individual. Apart from being rather beneficial for training deep neural networks, DP-SGD has been employed in various deep learning projects aiming at user privacy protection, including image classification and natural language processing chores [14].

Moreover, DP can be extended to additional parts of deep learning, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), by using methods comparable to those utilised in artificial neural networks. DP techniques can be applied, for photo categorisation, on CNN convolutional layers under training. This is done to make sure that personal privacy protection is not compromised by the gradients produced from particular training data. Similarly, DP methods are used into RNN backpropagation mechanism used for sequence prediction or natural language problems. This helps the research control privacy while yet allowing the learning from large databases. Included into these models, distributed processing (DP) helps the research to maintain privacy without appreciably compromising the performance of the model [15].

Challenges in Training Deep Models under DP

Even if differential privacy presents robust privacy protection, its application to deep learning models creates various difficulties. Among the most important problems is the tradeoff that has to be done between privacy and model performance. Although introducing noise to the gradients over the training process is necessary to protect users' privacy, doing so can lead to a lower model accuracy. Deep learning is a particularly challenging issue since often major results depend on great accuracy. A low ϵ value, that is, a low degree of noise, results in a more marked performance loss in the model. This is what is needed to guarantee high degree of privacy. the research of the most crucial challenges in the DP application process to deep learning models is to find the ideal balance between the incorporation of noise and the usability of the model [16].

The scalability of deep learning techniques using DP offers even another difficulty to be solved. Sometimes during deep neural network training processing large volumes of data and running millions of computations is necessary. Expanding these techniques to big datasets or models is more challenging since the insertion of noise into the research of these computations increases the memory use and the processing cost. Further difficult is the quest for suitable noise distribution and efficient scaling of it for large-scale training of deep learning. For example, depending on the size of the model, the size of the dataset, and the privacy budget, the quantity of noise must be continuously modified using DP-SGD. This seriously complicates the training process and requires careful customisation [17].

Another difficulty to be addressed is the generalisation potential of models learnt using DP. Deep learning models are supposed to effectively extend from training data to data not seen before, so the noise generated by DP could hamper this capacity to generalise. DP could make it more difficult for the model to grasp complicated data patterns via distortion of the gradients with noise. This can thus produce either overfitting or underfitting of the model. Regarding jobs needing a high degree of accuracy, such as autonomous driving or medical diagnostics, this is highly important since even little changes in model performance might have major consequences. Ensuring that the DP approaches have no meaningful effect on the generalising capacity of the model still presents a difficult task [18].

Thus, DP generates various difficulties in terms of model validity, scalability, and generality even if it offers a good framework for preserving personal privacy in deep learning. More research on optimising DP mechanisms for deep models, combining protection of users' privacy with maximising model performance, and developing new approaches to scale DP for big, complex networks is critically essential if we are to climb beyond these problems.

4. DP IN FEDERATED LEARNING

Federated Learning Overview

Federated learning is a decentralised paradigm used in machine learning. Under this approach, models are trained simultaneously across numerous devices or edge nodes and sensitive data is not centralised. Federated learning lets the model be trained locally on the research using the data saved on every device. This is not the case with the traditional method raw data is transferred to a central server for processing needs. After local training is finished, only the changes to the model, like gradients, are sent to the central server. Here the updates are compiled on the

central server into a worldwide model. With this approach, the research can teach machine learning models on a large volume of scattered data without breaching any data security or privacy [19].

Federated learning has gained significant attention for its ability to handle sensitive data, such as medical records, financial information, and personal data on mobile devices, while preserving privacy. Although advanced machine learning models are still evolving, federated learning helps address privacy concerns by keeping data localized and limiting model updates. However, ensuring privacy when aggregating model updates across multiple devices remains a challenge. One of the key techniques for protecting individual privacy in federated learning is differential privacy [20].

Privacy Mechanisms in Federated Learning

Privacy concerns arise in federated learning because model updates may unintentionally expose sensitive data. This poses risks, as adversaries could analyze aggregated updates to infer individual information. Since the model is trained across multiple devices, each holding potentially sensitive data, safeguarding privacy is crucial. To mitigate these risks, federated learning systems incorporate differential privacy and other privacy-preserving techniques.

Strongly related to DP, frequent privacy strategy applied in federated learning is incorporating noise to model updates. Every device produces noise to the computed gradients during training before the local model changes are sent to the central server. This takes place before the training session. Usually generated from a Gaussian or Laplace distribution, this noise guarantees that the aggregated statistical model is not substantially affected by individual data points. Moreover, considered in calibration of the degree of noise are the intended level of privacy (ϵ) and the parameter sensitivity of the model. However, federated learning can use secure aggregation to guarantee that, even in the instance that any of the model updates are stolen, the privacy of the devices involved in the process is not infringed [21].

Homomorphic encryption and secure multi-party computation (SMPC) are two powerful techniques that, when combined with differential privacy, further enhance the security and privacy of federated learning systems. These methods enable a central server to aggregate model updates without directly accessing or reviewing individual data, preventing the exposure of sensitive information. By ensuring that raw data remains encrypted or distributed across multiple parties, they provide strong privacy safeguards. These techniques are particularly valuable in highly sensitive fields, such as medical data analysis and financial transactions, where privacy is critical. While they offer an additional layer of security, they often come with increased computational demands, requiring efficient implementation to balance privacy with performance [22].

How DP Protects Federated Data

Differential privacy is one of the most powerful techniques for safeguarding user privacy in federated learning. It enhances security by adding carefully calibrated noise to local model updates before they are transmitted to the central server. Federated learning, a specialized area of machine learning, enables multiple devices to collaboratively train a model without sharing their raw data. By integrating DP, even if an adversary gains access to the aggregated updates, they will be unable to extract precise or sensitive information about individual users. The level of noise introduced is determined by the model's sensitivity and the privacy budget (ϵ), ensuring a balance between maintaining strong privacy protection and enabling effective model training.

Under the context of federated learning with DP, the privacy budget is under control over numerous training runs. Since every cycle of federated learning comprises numerous devices contributing their model updates, DP ensures the total impact of the changes does not exceed the privacy limit. However, the full privacy guarantee is under control by the overall noise produced throughout all rounds, which the budget for privacy sets. The device creates noise at every round, but this level governs the general privacy protection. When numerous rounds of model updates are under review, this cumulative noise approach serves to lower the capacity of an adversary to reverse-engineer crucial information [23]. This is true even in cases of multiple rounds of model upgrades.

Moreover, federated learning can leverage Local Differential Privacy (LDP) variance of DP. This local noise addition offers an additional degree of privacy since the central server can only review the noisy copies of the updates, not

access the raw updates. Sometimes widely scattered devices might not fully trust the central server, like in mobile apps or Internet of Things (IoT) networks [24]. These kinds of situations demand precisely this kind of approach.

DP in federated learning helps to protect personal data from public access while the central server can continue to compile pertinent updates to enhance the overall model. A prominent strategy for privacy-preserving machine learning is combining DP with federated learning. This method ensure that personal data is never exposed while still enabling the development of robust and accurate models from many sources.

5. KEY PROBABILITY DISTRIBUTIONS IN DP

Gaussian Distribution

Out of the many probability distributions discussed for differential privacy, the Gaussian distribution also known as the normal distribution is one of the most popular. It is mainly used in the noise injection process to achieve privacy in machine learning models. Differential privacy adds noise to the gradients or parameters of the model during training, which means that the contribution of a single data sample is reduced but the overall accuracy of the model is maintained. The Gaussian distribution is the most popular distribution due to its simplicity and mathematical analyzability and therefore, it is frequently employed for privacy in machine learning.

Gaussian noise enables the application of the Gaussian mechanism for differential privacy. The mean of this noise is zero and the variance is governed by the privacy budget (ϵ). This variance is a function of the sensitivity of the function being protected. To this end, the method ensures privacy by ensuring that no single data point drastically affects the model output. The amount of noise that must be added to a function increases with its sensitivity since the amount of noise is proportional to sensitivity. The Gaussian mechanism is most appropriate when there is a great need to achieve a correct trade-off between the privacy and accuracy of the model because the noise can be made to fit the calculation precisely [25]. This adaptability makes the Gaussian mechanism very effective.

Controlling the total privacy budget regulates the total privacy leakage as the Gaussian distribution is usable in differential privacy and is composable across multiple computations. This is one of the main reasons for employing the Gaussian distribution. Additionally, it has some other nice properties, for instance, the sub-Gaussian noise property. These properties are very important when dealing with continuous data and when the magnitude of privacy loss must be accurately determined to achieve a good trade-off between privacy and utility.

Laplace Distribution

Another used probability distribution in DP is the Laplace distribution. Techniques requiring more strict privacy guarantees find particular utility for this distribution. Usually, the Laplace mechanism is applied in situations when the function has rather great sensitivity. This also applies in handling discrete data. On the other hand, since its tails are thicker than those of the Gaussian distribution, the Laplace distribution is more sensitive to extreme outliers. The Gaussian distribution has noticeably narrower tails. Since it gives an opportunity to attain the target, this characteristic could be beneficial when the objective is to provide greater privacy safeguards under conditions when outliers in the data may otherwise expose too much about individual data points.

Noise emanating from a Laplace distribution is added to the Laplace mechanism. The mean of this distribution is zero; the desired degree of privacy (ϵ) and the sensitivity of the function help to determine the scaling coefficient. When exact findings are less important than privacy, as in data searches, the Laplace approach is frequently utilised [26]. This is so since the Laplace mechanism prioritises privacy above precision.

Laplace distribution in differential privacy is that it simplifies the mathematical analysis, making it easier to implement in real-world applications. The Laplace mechanism is particularly useful in federated learning and distributed systems, as it meets local differential privacy requirements. In such systems, noise is added directly to individual data points before aggregation, ensuring privacy at the local level while still allowing useful insights to be extracted from the data.

Other Distributions and Use Cases

Two most typically used distributions in differential privacy are the Gaussian and Laplace ones. Still, numerous different probability distributions can also be applied based on the particular privacy requirements and the characteristics of the data under control. These distributions offer flexibility for guarantees of privacy for certain applications of machine learning.

- a. **Exponential Distribution:** The exponential distribution is commonly used in modeling waiting times and skewed data because it naturally fits such scenarios. In differential privacy, it is particularly useful for adding noise in situations where events occur at a steady rate, such as count data or event-based data. As an extension of the Laplace mechanism, the exponential mechanism is designed for non-symmetric data. It offers another way to introduce noise while preserving privacy, making it a valuable tool for handling sensitive information in various data-driven applications [27].
- b. **Binomial Distribution:** The binomial distribution is well-suited for discrete data with binary outcomes, such as success or failure. In differential privacy, it can be used in mechanisms designed for tasks like classification or binary outcome prediction, ensuring that the overall model remains accurate while protecting individual data privacy. For example, in healthcare data analysis, the binomial approach can help safeguard patient privacy while still identifying patterns in sensitive data. If the outcomes are binary—such as whether a person has an illness or not—this method allows useful insights to be extracted without exposing personal information [28].
- c. **Geometric Distribution:** The geometric distribution models the number of trials needed before a successful outcome occurs in a series of independent Bernoulli experiments. In the context of differential privacy, it is particularly useful for failure-time data or count-based statistics. This distribution ensures privacy while allowing researchers to focus on counting occurrences until a specific event or outcome is reached. It provides a practical approach for releasing noisy data while maintaining privacy, making it ideal for situations where the goal is to track events or outcomes over time without exposing sensitive information. [29].
- d. **Poisson Distribution:** The Poisson distribution is widely used to model the frequency of events occurring within a specific time period or area. In fields like traffic monitoring or telecommunications, it is especially useful for handling rare events or small volumes of data. In the context of differential privacy, the Poisson process can be applied to introduce noise into collected data, particularly in environments where events are infrequent and the data may not be evenly distributed. This ensures privacy while allowing for accurate analysis in situations where occurrences are sparse or irregular [30].

The choice of distribution in differential privacy depends on the specific application, the type of data, and privacy requirements. Each distribution has its own characteristics and trade-offs, so the selection of the most appropriate one is influenced by these factors. In the context of deep learning, federated learning, and other machine learning applications, the right distribution can help maintain valuable utility while also protecting user privacy [31]. By carefully selecting the distribution, researchers can achieve a balance that ensures both privacy and effective performance.

6. COMPARING DP VARIANTS

Differentially Private Stochastic Gradient Descent

One of the most commonly used optimization techniques for training machine learning models with differential privacy is Differentially Private Stochastic Gradient Descent. This method of using gradient descent with perturbed gradients to train models is an extension of the standard stochastic gradient descent approach. DP-SGD works by first clipping the gradients of each mini-batch to a predefined norm. Then noise from a Gaussian distribution is added to the clipped gradients and the model parameters are updated. This guarantees that no single data point impacts the model significantly, thus protecting the privacy of individuals.

DP-SGD is typically formulated with two key parameters to describe the privacy guarantee of the algorithm: ϵ , the privacy budget, which indicates the level of privacy; and δ (delta), which is the probability of a privacy breach. DP-SGD is a strong privacy preserving training method and can be used to train deep learning models on sensitive data

including healthcare, financial, and personal data. However, the noise added to the gradients increases the cost, and thus there is a trade-off between privacy and accuracy. With complex datasets or high dimensional models, these trade-offs may become more apparent, leading to a decline in the accuracy of the model [32].

DP-SGD has two major advantages: It has a very sound theoretical basis and it is easy to implement. It can be used on a wide range of models and datasets, and is suitable for a variety of machine learning tasks. However, there is some issue with the need for much hyperparameter tuning to find the right balance between privacy and utility. Moreover, the noise addition and gradient clipping steps are computationally expensive and often lead to longer training times.

DP-Adam and Other Optimizers

DP-Adam is an extension of the popular Adam optimizer, which has been further improved to fulfill the DP requirements during the optimization process. The original Adam optimizer adapts the learning rate for each parameter of the model based on the first and second moments of the gradients, the mean and variance, respectively. Like DP-SGD, DP-Adam also adds noise to the updates and clips the gradients to a certain norm. However, the contribution of DP-Adam is especially significant in the learning phase since it incorporates Adam's adaptive learning rate policy that assists in finding a good trade-off between privacy and optimization.

DP-Adam is most effective when training deep neural networks, and while Adam's flexible step-sizes can help converge, it also makes the network sensitive to hyperparameter adjustments and noisy gradients. This is a major advantage of DP-Adam over DP-SGD, in that it is able to converge faster and perform better, especially for large models where the noise injected by DP-SGD can hinder training [33]. This is especially the case with large models that have many parameters. Furthermore, DP-Adam can better control the gradient updates with respect to changing gradient norms during the training process compared to a fixed learning rate. Although there are these benefits of DP-Adam, it is still similar to DP-SGD in many aspects such as the privacy-accuracy trade off. The noise introduced in DP-Adam can still lead to severe accuracy degradation, particularly in critical applications where the noise level needed for privacy is relatively high.

However, DP-Adam is more complicated than DP-SGD and has some hyperparameters that are not easy to set, which will increase the difficulty of training and applying models. In addition, other optimisers including DP-SAM Grad and DP-LARS are variants that are proposed to address specific issues in deep learning. Some of the other challenges include managing large datasets and improving convergence. On the other hand, because of the complexity of the algorithms and the absence of studies on their performance in the privacy environment, these optimisers are not as popular as DP-SGD and DP-Adam [34].

7. THEORETICAL AND PRACTICAL GAPS IN DP

Theory vs. Real-World Applications

DP offers rigid theoretical assurances for individual privacy in datasets or machine learning algorithms. This method ensures that the result of a query or model cannot be much changed by adding or removing the research data point. Although DP provides guarantees theoretically, the actual efficiency of the DP systems applied in the real world still lags far behind. From a theoretical standpoint, it is possible to construct DP systems such that they provide a strong privacy protection by use of noise in a way that hides significant personal data about individuals. Still, given the inherent complexity of the data and models used in the real world, these systems sometimes fail to live up to expectations when applied in pragmatic environments.

Datasets can be exceedingly noisy, unstructured, or high-dimensional, hence DP can be difficult to employ practically. This causes the system problems. Real-world applications can demand for the management of complex and heterogeneous data (such as images, text, and sensor data); the process of adapting domain-specific programming to such data formats is not always simple. Further challenging the preservation of one's privacy and the planned model performance is the interaction of privacy limitations with pragmatic objectives including the necessity for quick processing and great accuracy [35].

Further, it is crucial to emphasise that although DP provides mathematical definitions of privacy budgets (ϵ) and failure probability (δ), these values are not always easily interpretable in respect to the real-world privacy challenges addressed. DP's conceptual roots are based on the assumption that everyone has privacy concerns. Meanwhile, this presumption could not always coincide with the privacy issues of the actual world. The research of the most crucial issues arising from trying to implement DP in real-world systems is the difference between theoretical models and actual needs.

Accuracy vs. Privacy Trade-offs

Managing privacy against accuracy is the research of the most important challenges DP presents. Differential privacy is preserved by including noise into the data or gradients during training of the model. Calculated by the parameters of ϵ and δ , the accuracy of the model directly relies on the noise level needed to obtain a specific degree of privacy. Higher degrees of privacy (lower ϵ) require more noise, so the model performs poorer even if this is essential. Deep learning models clearly show this trade-off since their great dimensionality and great parameter space improve their sensitivity to noise.

Regarding applications used in the real world, this trade-off becomes rather important since data owners usually want to obtain high utility, that is, accurate forecasts or insights, while nevertheless maintaining their privacy. Privacy guarantees are extremely important in sensitive industries like finance and healthcare, where the data used is frequently rather sensitive and routinely contains personal information. Conversely, excessive noise could lead to a declining accuracy of the model to an unwelcome level, therefore reducing its value in real-world decision-making [36].

Adaptive noise addition and hybrid approaches combining DP with other techniques, such as federated learning, two instances of the systems researchers have devised to compromise accuracy and privacy obligations. Conversely, the achievement of an ideal balance is still a subject of discussion and more study is needed to design better systems that give major privacy assurances while preserving accuracy to the best degree possible.

Robustness in DP Mechanisms

The robustness of the DP mechanisms begs another significant problem of interest. Resilient DP mechanisms are those which can preserve their privacy assurances against numerous challenges like noisy data, model complexity, and adversarial attacks. On the other hand, most of the present DP approaches find it difficult to stay robust in such kinds of situations. Working with large-scale datasets or advanced deep learning models could produce the extra noise rising unrealistically high. This makes the method useless or ineffective in terms of maintaining a fair degree of privacy and yet obtaining good model performance.

The way gradient clipping and noise addition act in the presence of high-dimensional data or sparse gradients, both of which are prevalent in deep learning, is among the most crucial issues that has to be solved if we are to ensure resilience. These events can lead the DP approaches to fail in providing both privacy and accuracy in practical deployed applications. To aggravate the situation, theoretical guarantees on the longevity of DP techniques in environments with adversarial threats are lacking. Nowadays, machine learning models rely more and more on being sheltered from adversarial attacks, which involve the attacker skilfully changing the data or model. Sometimes the DP techniques could not be sufficient to provide enough defence against such attacks.

Moreover, occurrences in the actual world may provide dynamic obstacles that make it difficult to simultaneously guarantee durability and privacy. Regarding federated learning or remote settings, for instance, it could be difficult to maintain the DP process resilience over a wide spectrum of devices or data sources. This is so since many nodes could have different degrees of network connectivity, processing capacity, and data quality [37].

Therefore, the issue of establishing that DP is robust in practice under a broad spectrum of real-world conditions remains persistent even if DP has shown great theoretical growth in terms of providing privacy assurances. Novel solutions are required to overcome these constraints and guarantee that DP methods may effectively safeguard privacy while maintaining robust performance across a spectrum of environments.

Differential privacy provides strong theoretical privacy assurances in principle; yet, bridging the gap between theory and practice in the real world remains a great difficulty. More study in the subject of confirming the durability of DP approaches in surroundings hostile, complicated, and loud is needed since the trade-off between accuracy and privacy still causes tremendous worry. Theoretical improvements are very essential to bridge these gaps and make DP a more effective and successful instrument for machine learning that respects individuals privacy [38].

8. CONCLUSION AND FUTURE DIRECTIONS

Summary of Findings

This review article has offered a comprehensive overview of differential privacy, commonly called as differentiated privacy, inside the paradigm of deep learning and federated learning. DP has developed as a sensible model for preserving machine learning privacy and data analysis integrity. Recently it has become very popular and provides strong theoretical guarantees that sensitive information about individual data points is secured. DP enables the research to obtain valuable insights from data without infringing the privacy of people by use of methods such the injection of controlled noise to datasets or model parameters. Although DP has rather fundamental theoretical roots, its implementation in actual applications presents several pragmatic difficulties.

In the field of deep learning, the application of DP presents major challenges for the training of intricate models. Particularly for large-scale models with high-precision data, the injection of noise required to preserve privacy could have a significant impact on the simulation accuracy. Moreover, applying DP in federated learning systems presents a special method for privacy preservation. This method lets sensitive information be secured concurrently with models maybe trained remotely. On the other hand, the dynamic character of data and the heterogeneity of data among far-off devices typically compromise the privacy protections DP offers in such environments.

Moreover, the trade-off between privacy and model correctness remains a prominent topic of interest even although DP has developed various techniques including the incorporation of Gaussian and Laplace noise. Particularly with regard to the guarantee of robustness against the challenges of the real world, such as adversarial attacks and noisy data, the DP methods obviously demand considerable improvement. Further, theoretical concepts of DP often cannot satisfy the requirements of real, high-dimensional datasets and privacy concerns important for specific applications. This implies that if we wish to have better degree of change, DP methods have to be more flexible and adaptive.

Research Opportunities and Challenges in DP

The application of DP to deep and federated learning opens up some significant research opportunities and problems. The improvement of better DP tools that can address the trade-off between privacy and model accuracy is the research of the most crucial directions that research should follow. Current noise addition techniques have a tendency to obviously reduce model performance. This can particularly be challenging in fields including healthcare, banking, and autonomous systems, where great precision is quite relevant. Research on adaptive noise techniques or hybrid privacy preserving systems that combine DP with other techniques such as homomorphic encryption or safe multiparty computation could be useful in reducing the effects of these concerns. Another interesting subject of the future research is the enhancement of the DP methods' robustness. This is because of the need to deal with noisy, high dimensional datasets and adversarial attacks. Strong DP methods will be essential to guarantee that privacy protections are maintained even when there are very powerful attackers or in a very complex data environment. Furthermore, given the increasing adoption of federated learning in decentralized systems, the improvement of the privacy protections offered by DP in federated learning settings is a key focus of research. Solutions that enable DP methods to extend their coverage to multiple distributed nodes without jeopardizing performance or privacy are required. Moreover, a big question still exists on the gap between theory and practice. While DP offers strong mathematical guarantees, these theoretical models often do not capture the richness and diversity of the data seen in the real world. This gap can be closed by developing DP models that are more context aware and adaptable. These models should be unpredictable to meet the different privacy needs of various applications. Furthermore, more attention should be paid to how practically the research interprets privacy budgets (ϵ and δ). This is important in sensitive areas like healthcare and banking where the problem is rather frequent. An area that remains relatively unexplored is the integration of DP with emerging machine learning paradigms. Some of these paradigms include

reinforcement learning and few shot learning, which pose specific privacy risks due to limited data or continuous learning operations.

REFERENCES

- [1] Ghazi, B., Golowich, N., Kumar, R., Manurangsi, P., & Zhang, C. (2021). Deep learning with label differential privacy. *Advances in neural information processing systems*, 34, 27131-27145.
- [2] Ziller, A., Usynin, D., Braren, R., Makowski, M., Rueckert, D., & Kaissis, G. (2021). Medical imaging deep learning with differential privacy. *Scientific Reports*, 11(1), 13524.
- [3] Wang, Y., Wang, Q., Zhao, L., & Wang, C. (2023). Differential privacy in deep learning: Privacy and beyond. *Future Generation Computer Systems*, 148, 408-424.
- [4] Papernot, N., Thakurta, A., Song, S., Chien, S., & Erlingsson, Ú. (2021, May). Tempered sigmoid activations for deep learning with differential privacy. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 10, pp. 9312-9321).
- [5] Bu, Z., Wang, H., Dai, Z., & Long, Q. (2023). On the convergence and calibration of deep learning with differential privacy. *Transactions on machine learning research*, 2023.
- [6] El Ouadrhiri, A., & Abdelhadi, A. (2022). Differential privacy for deep and federated learning: A survey. *IEEE access*, 10, 22359-22380.
- [7] Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., & Muralidhar, K. (2022). A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Computing Surveys*, 55(8), 1-16.
- [8] Yousefpour, A., Shilov, I., Sablayrolles, A., Testuggine, D., Prasad, K., Malek, M., ... & Mironov, I. (2021). Opacus: User-friendly differential privacy library in PyTorch. *arXiv preprint arXiv:2109.12298*.
- [9] Gutiérrez, N., Otero, B., Rodríguez, E., Utrera, G., Mus, S., & Canal, R. (2024). A Differential Privacy protection-based federated deep learning framework to fog-embedded architectures. *Engineering Applications of Artificial Intelligence*, 130, 107689.
- [10] Sharma, J., Kim, D., Lee, A., & Seo, D. (2021). On differential privacy-based framework for enhancing user data privacy in mobile edge computing environment. *IEEE Access*, 9, 38107-38118.
- [11] Gwon, H., Ahn, I., Kim, Y., Kang, H. J., Seo, H., Choi, H., ... & Kim, Y. H. (2024). LDP-GAN: Generative adversarial networks with local differential privacy for patient medical records synthesis. *Computers in Biology and Medicine*, 168, 107738.
- [12] Cynthia Dwork and Aaron Roth (2014), *The Algorithmic Foundations of Differential Privacy*, Foundations and Trends in Theoretical Computer Science
- [13] Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, Li Zhang (2016). *Deep Learning with Differential Privacy* arXiv:1607.00133
- [14] Jiang, B., Li, J., Yue, G., & Song, H. (2021). Differential privacy for industrial internet of things: Opportunities, applications, and challenges. *IEEE Internet of Things Journal*, 8(13), 10430-10451.
- [15] Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. *Scientific reports*, 12(1), 1953.
- [16] Domingo-Ferrer, J., Sánchez, D., & Blanco-Justicia, A. (2021). The limits of differential privacy (and its misuse in data release and machine learning). *Communications of the ACM*, 64(7), 33-35.
- [17] Luo, Z., Wu, D. J., Adeli, E., & Fei-Fei, L. (2021). Scalable differential privacy with sparse network finetuning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5059-5068).
- [18] Li, Y., Duan, Y., Maamar, Z., Che, H., Spulber, A. B., & Fuentes, S. (2021). Swarm Differential Privacy for Purpose-Driven Data-Information-Knowledge-Wisdom Architecture. *Mobile Information Systems*, 2021(1), 6671628.
- [19] Guo, C., Karrer, B., Chaudhuri, K., & van der Maaten, L. (2022, June). Bounding training data reconstruction in private (deep) learning. In *International Conference on Machine Learning* (pp. 8056-8071). PMLR.
- [20] Ponomareva, N., Hazimeh, H., Kurakin, A., Xu, Z., Denison, C., McMahan, H. B., ... & Thakurta, A. G. (2023). How to dp-fy ml: A practical guide to machine learning with differential privacy. *Journal of Artificial Intelligence Research*, 77, 1113-1201.

- [21] Yan, C., Yan, H., Liang, W., Yin, M., Luo, H., & Luo, J. (2024). DP-SSLoRA: a privacy-preserving medical classification model combining differential privacy with self-supervised low-rank adaptation. *Computers in Biology and Medicine*, 179, 108792.
- [22] Hu, H., Han, Q., Ma, Z., Yan, Y., Xiong, Z., Jiang, L., & Zhang, Y. (2023, October). PV-PATE: An Improved PATE for Deep Learning with Differential Privacy in Trusted Industrial Data Matrix. In *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data* (pp. 477-491). Singapore: Springer Nature Singapore.
- [23] Farayola, O. A., Olorunfemi, O. L., & Shoetan, P. O. (2024). Data privacy and security in it: a review of techniques and challenges. *Computer Science & IT Research Journal*, 5(3), 606-615.
- [24] Yang, M., Guo, T., Zhu, T., Tjuawinata, I., Zhao, J., & Lam, K. Y. (2024). Local differential privacy and its applications: A comprehensive survey. *Computer Standards & Interfaces*, 89, 103827.
- [25] Oyewole, A. T., Oguejiofor, B. B., Eneh, N. E., Akpuokwe, C. U., & Bakare, S. S. (2024). Data privacy laws and their impact on financial technology companies: a review. *Computer Science & IT Research Journal*, 5(3), 628-650.
- [26] Yan, B., Li, K., Xu, M., Dong, Y., Zhang, Y., Ren, Z., & Cheng, X. (2024). On protecting the data privacy of large language models (llms): A survey. *arXiv preprint arXiv:2403.05156*.
- [27] Olabim, M., Greenfield, A., & Barlow, A. (2024). A differential privacy-based approach for mitigating data theft in ransomware attacks. *Authorea Preprints*.
- [28] Yang, L., Tian, M., Xin, D., Cheng, Q., & Zheng, J. (2024). AI-Driven Anonymization: Protecting Personal Data Privacy While Leveraging Machine Learning. *arXiv preprint arXiv:2402.17191*.
- [29] El Mestari, S. Z., Lenzini, G., & Demirci, H. (2024). Preserving data privacy in machine learning systems. *Computers & Security*, 137, 103605.
- [30] Seeman, J., & Susser, D. (2024). Between privacy and utility: On differential privacy in theory and practice. *ACM Journal on Responsible Computing*, 1(1), 1-18.
- [31] Yalamati, S. (2024). Data Privacy, Compliance, and Security in Cloud Computing for Finance. In *Practical Applications of Data Processing, Algorithms, and Modeling* (pp. 127-144). IGI Global.
- [32] Bakare, S. S., Adeniyi, A. O., Akpuokwe, C. U., & Eneh, N. E. (2024). Data privacy laws and compliance: a comparative review of the EU GDPR and USA regulations. *Computer Science & IT Research Journal*, 5(3), 528-543.
- [33] Thummisetti, B. S. P., & Atluri, H. (2024). Advancing healthcare informatics for empowering privacy and security through federated learning paradigms. *International Journal of Sustainable Development in Computing Science*, 6(1), 1-16.
- [34] Chukwunweike, J. N., Yussuf, M., Okusi, O., & Oluwatobi, T. (2024). The role of deep learning in ensuring privacy integrity and security: Applications in AI-driven cybersecurity solutions. *World Journal of Advanced Research and Reviews*, 23(2), 2550.
- [35] Chen, Y., & Esmailzadeh, P. (2024). Generative AI in medical practice: in-depth exploration of privacy and security challenges. *Journal of Medical Internet Research*, 26, e53008.
- [36] Tertulino, R., Antunes, N., & Morais, H. (2024). Privacy in electronic health records: a systematic mapping study. *Journal of Public Health*, 32(3), 435-454.
- [37] Zhu, M., Yuan, J., Wang, G., Xu, Z., & Wei, K. (2024). Enhancing Collaborative Machine Learning for Security and Privacy in Federated Learning. *Journal of Theory and Practice of Engineering Science*, 4(02), 74-82.
- [38] Li, Y., Yan, H., Huang, T., Pan, Z., Lai, J., Zhang, X., ... & Li, J. (2024). Model architecture level privacy leakage in neural networks. *Science China Information Sciences*, 67(3), 132101.