IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Privacy vs. Utility in Federated Learning: An Experimental Analysis of Noise Injection Techniques

N.R. LEOPE[1], J.H.P. ELOFF[1], and M.T. DLAMINI[12]

[1]Cyber Security Research Group (CySec), Department of Computer Science, EBIT Faculty, University of Pretoria, Pretoria, South Africa (e-mail: u19378816@tuks.co.za; jan.eloff@up.ac.za)

[2]Information and Cyber Security Centre, Defence & Security Cluster, CSIR, Pretoria, South Africa (e-mail: TDlamini1@csir.co.za)

Corresponding author: J. H. P. Eloff (e-mail: jan.eloff@up.ac.za).

ABSTRACT Federated Learning (FL) enables decentralized model training, while maintaining the privacy of the underlying individual datasets. Therefore, FL can resolve some intrinsically privacy-sensitive challenges in domains, such as healthcare and finance. However, privacy preservation usually comes with a trade-off on the usefulness (i.e., utility) of the information. The research problem is how to optimize this inversely proportional trade-off balance between privacy and utility. This study uses an experimental comparative analysis, in a synthetic healthcare setting, of different noise types (i.e., Gaussian, Laplacian, Poisson, Uniform, and Exponential) injected on the client side at the input-feature level prior to local training to enhance privacy in FL. We explore the impact of these noise types on the privacy–utility trade-off in FL data. The findings indicate that although Laplacian, Poisson, and Exponential types of noise provides stronger obfuscation which often comes at the cost of utility. This confirms and amplifies the trade-off in maintaining the usefulness of the data against its privacy. More importantly, the findings also show that Gaussian noise generally offers the best trade-off between privacy and utility on this task, suggesting a practical default for privacy-aware FL in healthcare-like environments.

INDEX TERMS Data Poisoning, Differential Privacy, Federated Learning, Malicious Data Injection, Privacy-Utility Trade-off

I. INTRODUCTION

DATA privacy breaches continue to leak patient health data to unauthorized threat actors, posing significant risks to the heavily regulated healthcare sector [1]. Privacy-preserving control measures are essential for safeguarding patient health data held by healthcare facilities. This is more critical for healthcare facilities that share and exchange patient health data to facilitate patient mobility, patient-centric healthcare services, referrals and continuous care. For instance, when parents move between provinces or states, uninterrupted access to their medical history is crucial to provide accurate prescriptions and continuous care. This requires visibility of patient healthcare data across multiple facilities.

At a strategic level, aggregating healthcare data from multiple facilities supports the training of artificial intelligence (AI) models that aid health ministries in responding to disease outbreaks. This became evident during the COVID-19 pandemic, when decision makers monitored disease outbreaks to enforce evidence-based travel restrictions and lockdowns. However, aggregating patient health data raises the need to preserve its privacy. This has become a key research area in FL.

FL has emerged as a promising paradigm that enables a decentralized model training whilst ensuring that the data remain local, thereby addressing privacy concerns in sensitive healthcare applications. FL is defined by McMahan *et al.* [2], as a type of machine learning technique in which several devices or users work together to train a common model, but each one stores its data locally instead of sharing it. This definition focuses on the shared model training and localized data. Bharati *et al.* [3] defined FL as a system in which a central agent coordinates the efforts of several clients to solve complex challenges using machine learning approaches while ensuring that data remain localized and private. Similarly, the definition in [3] focused on a centralized aggregator, which is synonymous with a shared model. However, the definition in [3] adds that over and above data localization, data must also remain private and not be exposed to untrusted third parties.

Kairouz *et al.* [4] defined FL as "a machine learning setting whereby multiple decentralized devices or organizations collaboratively train a model under privacy constraints, distinguishing it from traditional centralized learning paradigms" . This definition is similar to that of [2] with respect to the shared model. Therefore, this study defines FL as a decentralized machine learning framework that facilitates collaborative model training across multiple clients while preserving data privacy. This definition is unique in its emphasis on both the structural decentralization and the privacy-preserving intent of FL.

Although FL introduces compelling benefits, especially in domains with strict data-sharing constraints, it is also susceptible to several well-documented threats, including *data poisoning attacks*, where malicious clients manipulate local data to corrupt the global model; *model inversion attacks*, where adversaries attempt to reconstruct original training data from shared gradients or updates; and *privacy breaches*, which may occur through side-channel attacks or unintended leakage in model updates. These vulnerabilities are captured in existing literature. For example, the work of [5], [6], and [7], illustrate how FL's distributed nature can become a double-edged sword—enhancing privacy superficially while opening new vectors for privacy breaches. Privacy breaches in FL refer to situations in which an adversary, possibly a participating client, infers sensitive information about another client's data without directly accessing it.

It should be noted that preserving privacy intrinsically reduces the usefulness (utility) of the data i.e. privacy and utility are inversely proportional. This paper hypothesizes that as noise is applied to provide stronger privacy guarantees, the utility of the model outputs typically degrades and vice versa. Therefore, this study is grounded in and extends existing FL literature by investigating the use of noise to provide strong privacy guarantees balanced with marginal effects on the utility of the data.

Accordingly, our primary research question is: *How do various noise types (such as Gaussian, Laplacian, Exponential, Uniform, and Poisson) differ in their balance of privacy versus data utility?* We hypothesize that each noise distribution will occupy a distinct point along the privacy–utility curve, offering varying privacy–utility trade-offs. A secondary question is: *Which noise mechanism achieves the best privacy–utility trade-off under realistic FL conditions?*

The selected noise types, i.e. Gaussian, Laplacian, Exponential, Poisson, and Uniform were chosen because of their prevalence in highly cited privacy-preserving machine learning research. The Gaussian and Laplace mechanisms are canonical in differential privacy and have been widely adopted in large-scale machine learning [8], [9].

More recently, Zhao and Wang [10] introduced a prefix-tree approach tailored to trajectory data that improved both privacy and utility outcomes. Their findings underscore the growing emphasis on domain-sensitive noise injection, which is aligned with the goal of this study.

The remainder of this paper is organized as follows. Section II outlines the key background concepts. Section III discusses related work on FL vulnerabilities and mitigation strategies. Section IV presents healthcare use cases. Section V describes the FL pipeline and noise injection points. Section VI details the proposed noise-based privacy framework. Sections IX and X present the experimental setup and the results, respectively. Section XI discusses the findings, and Section XII concludes the paper and highlights future directions of research in FL.

II. BACKGROUND

FEDERATED Learning (FL) enables collaborative model training across decentralized devices while preserving the data locality. However, security and privacy challenges persist owing to adversarial threats and privacy risks. This section outlines key concerns, including Byzantine attacks, Secure Multi-Party Computation (SMPC), and Differential Privacy (DP), to contextualize FL security research. Although this study does not focus on Byzantine fault mitigation or SMPC, understanding these aspects remains crucial for designing privacy-aware FL architectures that effectively balance the trade-off between data privacy and utility.

A. BYZANTINE FAULTS IN FEDERATED LEARNING

Byzantine faults in FL occur when one or more participating clients maliciously behave. For example, injecting incorrect or poisoned gradients, submitting random noise, or colluding with others can deliberately distort a global model, reduce its accuracy, or extract sensitive information. This is particularly critical in high-stakes applications such as healthcare and finance [5]. To counteract these threats, Byzantine-resilient aggregation methods such as coordinate-wise median, trimmed mean, and Krum have been developed. In addition, anomaly-detection techniques that leverage trust-based heuristics and reinforcement learning aim to identify adversarial updates [11].

1) Byzantine-Resilient Aggregation Techniques

Byzantine-resilient aggregation methods have been designed to mitigate the impact of malicious or faulty clients on FL by robustly aggregating model updates. Notable techniques include coordinate-wise median, trimmed mean, and Krum. The coordinate-wise median computes the median of each model parameter across clients to reduce the influence of outliers [12]. The trimmed mean method removes a fixed percentage of the highest and lowest values of each parameter before averaging. This approach mitigates the effects of extreme and adversarial values [13]. Finally, the Krum method selects the single update closest to the Euclidean distance to the majority of other updates, effectively filtering out anomalies and adversarial contributions. All of these methods provide statistical robustness and are foundational to many defense strategies against poisoning and Byzantine attacks in FL.

2) Anomaly Detection in Healthcare Federated Learning

Beyond robust aggregation, anomaly-detection techniques play a crucial role in proactively identifying adversarial behavior in FL, particularly in healthcare, where model integrity, confidentiality, and data sensitivity are paramount. Recent studies investigated audit- and graph-based trust mechanisms. Audit-based trust mechanisms include TrustFed [14], which proposes HiAudit-FL, a two-tier auditing framework that alternates between lightweight checks and deeper inspections to identify malicious clients. The authors cast the auditing decision process as a Partially Observable Markov Decision Process (POMDP), and then used deep reinforcement learning to determine the best trade-off between detection accuracy and auditing overhead. For graph-based detection, G^{2}uardFL [15] constructed graphs of client behavior to detect structural anomalies and isolate backdoor attacks based on graph clustering techniques tailored to healthcare-specific training tasks.

These approaches complement aggregation defenses by providing intelligent detection mechanisms suited to complex and sensitive environments such as healthcare FL. Although Byzantine-resilient aggregation techniques enhance robustness against malicious updates, they primarily address reliability and correctness and not privacy. These methods assume a threat model that focuses on poisoning, rather than inference attacks. Therefore, they do not protect against privacy leakage from gradient updates, which is central to the present study. Furthermore, robust aggregation can degrade data utility, complicating the privacy–utility trade-off in healthcare applications in which diagnostic accuracy is critical.

B. PRIVACY-PRESERVING MECHANISMS: SECURE MULTI-PARTY COMPUTATION AND DIFFERENTIAL PRIVACY

Although FL keeps raw data local, adversaries can still infer sensitive information from model updates through techniques such as *membership inference attacks*—which determine whether a particular data point was part of a client's training set—and *gradient inversion attacks*, which reconstruct input data from shared gradients [16]–[18]. SMPC enhances privacy by enabling secure aggregation through techniques such as Shamir's secret sharing, homomorphic encryption, and garbled circuits, which prevent direct access to individual model updates [19], [20].

DP provides formal, quantifiable guarantees by adding carefully calibrated random noise to model updates, thereby limiting the influence of any single data point on the output. This ensures that an adversary cannot confidently determine whether an individual's data are included in the training set, thereby protecting against re-identification [9]. Gaussian and Laplacian noise perturbations help balance privacy and utility, whereas adaptive noise dynamically scales to adjust noise levels based on model sensitivity. Hybrid approaches combining DP and SMPC further strengthen the privacy and security of FL [20], [21].

The SMPC provides strong cryptographic security for aggregating model updates without revealing individual contributions. However, it incurs substantial computational and communication overheads, making it less practical for bandwidth-constrained healthcare environments. More importantly, while SMPC secures the aggregation process, it does not obfuscate the gradients themselves, leaving the system vulnerable to inference attacks if other protections such as DP are not employed. Thus, SMPC alone does not satisfy the dual goals of preserving privacy and maintaining utility in noisy healthcare-data scenarios.

1) Bridging Security, Privacy, and FL Resilience

A resilient FL system must defend against unreliable participants, protect individual data privacy, and prevent eavesdropping on the model updates. Byzantine fault tolerance is first achieved using robust aggregation rules, such as Krum, which detect and counteract malicious or faulty client updates, ensuring that the global model continues to converge even if some participants send poisoned gradients [13]. DP then adds a calibrated layer of protection by injecting random noise into client updates (or, in other implementations, the final model). Thus, by bounding the information, an adversary can infer any single data point and guard against membership inference or reconstruction attacks [9]. Finally, SMPC protocols such as those mentioned in Section II, cryptographically split each client's update such that the central server sees only an encrypted sum. This prevents it (or any other colluding party) from recovering individual contributions during aggregation [22]. Together, these three defenses form a complementary triad that addresses the key threats in FL: unreliable or malicious participants, privacy leaks from shared information, and exposure to raw model updates.

This study examined the impact of noise perturbation mechanisms on FL, focusing on their effects on convergence, privacy guarantees, and system reliability. Although extensive research has been conducted on enhancing FL security through Byzantine fault-tolerant aggregation, SMPC-based secure computations, and DP mechanisms, significant challenges remain in optimizing the privacy–utility trade-off. Many existing studies either focused on robust aggregation to mitigate Byzantine faults or explored privacy-preserving techniques, such as DP and SMPC, in isolation, often neglecting their combined impact on model performance and privacy. Furthermore, current approaches primarily evaluate privacy guarantees without thoroughly examining how different noise distributions affect the model convergence, accuracy, and overall system robustness in adversarial settings. This gap highlights the need for a deeper investigation of the interplay between noise-perturbation strategies and FL security. Addressing this challenge is crucial for designing FL architectures in order to achieve strong privacy guarantees and enhanced data utility, and thus also reliable model performance.

DP is well suited to limit data leakage, but introducing noise directly affects model performance. Even small reductions in utility can lead to clinically significant errors in healthcare. Moreover, many DP implementations in FL

assume homogeneous data distributions and overlook the unique challenges of non-IID and sparse medical data. This raises concerns about fairness and generalizability. Therefore, evaluating diverse noise mechanisms in a privacy–utility context, as in this study, is essential for optimizing real-world deployment of privacy-preserving FL in healthcare. To achieve this, it is important to establish a formal definition of the core concepts of data utility and privacy.

C. FORMAL DEFINITIONS OF UTILITY AND PRIVACY

In this study, **Utility** is defined as the degree to which a privacy-preserving transformation retains the essential information and patterns of the original dataset required for accurate inference. In other words, high utility means that noise-injected (synthetically perturbed) healthcare data still preserves key statistical characteristics or trends, enabling downstream models to yield similar insights as they would on raw data. Utility is conceptually inversely proportional to privacy, and stronger privacy protection typically degrades the usefulness of data.

Privacy refers to the degree to which a transformed dataset protects sensitive information from original data through reconstruction or inference by adversaries. High privacy implies that noise-injected data significantly obfuscates individual data points and identifiable patterns, thereby preventing effective reverse engineering. Privacy increases as the transformed data become less representative of the original input, thereby decreasing recoverability.

III. RELATED WORK

THE concept of FL was introduced into the limelight by Google in 2016 as an approach to address privacy concerns in machine learning [2]. Since then, FL has attracted considerable attention across multiple domains. For example, in the healthcare sector, Hu *et al.* [23] used FL to detect membership inference attacks and prevent medical privacy breaches. Early research largely focused on basic implementations and exploring the potential of FL. Gradually, the focus shifted to addressing inherent privacy weaknesses. Although FL has been applied in different contexts to help reduce data privacy breaches, significant challenges related to model vulnerabilities remain. Liu *et al.* [24] also agreed that FL continues to suffer from widespread data-privacy breaches. This may be due to the susceptibility of the shared or centralized model to inadvertently leak sensitive patient health data. Liu *et al.* [24] argued that these vulnerabilities are partly due to the decentralized nature of FL, which raises privacy concerns regarding data poisoning, model inversion, and Byzantine faults. This underscores the increasing need to propose more privacy-preserving solutions for FL to ensure that patient health data are not leaked to threat actors.

A. DIFFERENTIAL PRIVACY IN FL

Recently, the privacy research community has increasingly explored Differential Privacy (DP) to provide strong privacy guarantees in FL. Zhao *et al.* [25] introduced the SecProbe framework, which integrates DP with robust aggregation methods to mitigate the effects of malicious and unreliable participants. Although this approach improves privacy, it can also degrade model accuracy if the added noise is not carefully calibrated.

The introduction of noise into data or model updates is a foundational technique for achieving DP, which aims to limit the risk of re-identification from shared outputs. The seminal work by Dwork *et al.* [9] introduced the formal concept of DP, demonstrating that adding Laplace or Gaussian noise can ensure that outputs do not compromise individual data points. More recently, Abadi *et al.* [8] presented the Moments Accountant method for rigorous privacy accounting during training, and Geyer *et al.* [26] applied DP in practical FL settings with real-world mobile data. Knowledge distillation has also been combined with FL to produce lightweight and explainable intrusion detection systems for resource-constrained devices, highlighting the trade-off between efficiency, interpretability, and privacy in applied FL [27].

In this study, we define *noisy data* as data perturbed by a calibrated stochastic mechanism, typically via additive noise, to reduce privacy leakage during training or aggregation while retaining acceptable utility. Understanding the influence of different noise distributions on this balance is central to our study. However, these mechanisms often face challenges in balancing privacy and utility, particularly in high-dimensional data scenarios. Therefore, this study extends the work of Dwork *et al.* [9] by investigating the effects of adding noise to FL to provide strong privacy guarantees balanced with optimal utility. Other researchers continue to explore external noise mechanisms and adaptive privacy budgets to improve this trade-off [28]. However, ensuring robust privacy guarantees without compromising utility remains a significant challenge in applying DP to FL.

B. LIMITATIONS OF EXISTING APPROACHES

The SMPC and DP techniques used in FL differ in their approaches to preserving the privacy of aggregated datasets. Although both are foundational to privacy-preserving FL, they have notable limitations. SMPC offers strong cryptographic guarantees but incurs high computational and communication overheads, making it less feasible for large-scale or resource-constrained settings [22]. DP provides formal guarantees but often relies on fixed noise addition strategies that may not adapt well to heterogeneous data distributions, particularly in sensitive domains such as healthcare [26].

Therefore, this study investigates a range of noise distributions—including Laplace, Gaussian, Uniform, Exponential, and Poisson distributions—as a flexible and empirically grounded approach to evaluating their privacy–utility trade-offs in FL. The goal is to identify noise models that offer both strong privacy protection and acceptable utility in real-world healthcare applications. While FL decentralizes participation and enables collaboration, it also suffers from challenges such as Byzantine faults and non-optimal participant selection, which impact model robustness and efficiency [11].

IV. USE CASES AND THREAT MODEL

THIS study examined how different types of noise, such as Gaussian, Poisson, Uniform, Exponential, and Laplacian, can enhance patient healthcare data privacy while simultaneously being cognizant of data utility in real-world healthcare scenarios. The case study began with an individual health facility in a provincial and national network of healthcare facilities.

Noise must be consistently applied to maintain the stability and utility of the model. Several types of noise arising from different institutions can make model aggregation unreliable and thereby reduce utility. Standardizing the noise injection across all stages of the FL process—such as during client-side training, communication, and server-side aggregation—can help ensure that the resulting global models remain consistent in structure and behavior. This consistency minimizes the variance in update distributions across clients, which in turn enhances overall utility and stability while preserving privacy guarantees. To analyze the impact of noise on the privacy–utility trade-off, we conducted a series of simulated empirical studies in different scenarios where the same type of noise was injected. These simulations suggest solutions to healthcare applications that balance privacy and utility.
To keep the main text concise the detailed use-case diagrams (facility–province–national) are provided in Appendix A.

1) Scope and Threat Model

We consider an honest but curious (passive) server that orchestrates training and may inspect model updates. The attacker can observe server-side aggregated updates and per-round metrics, but cannot access raw client data, client identities, or private keys. Network-layer and side channel attacks are out of scope. Unless stated otherwise, clients follow the protocol (data remain local; no raw samples are shared), and privacy risk arises from information that may be inferred from shared model updates. For experiments using central/example-level DP–SGD with Gaussian noise and RDP accounting, we claim a formal (ϵ, δ) guarantee against any adversary observing the DP mechanism's outputs (including the server); for experiments using Gaussian Local DP (LDP), features are privatized on-device before sharing, so the guarantee holds even if the server is fully untrusted, by post-processing.By contrast, our noise ablations (e.g., uncalibrated feature perturbations) are reported only as privacy–utility *proxies* and are not formal privacy guarantees.

V. FEDERATED LEARNING PIPELINE

THE decentralized FL process begins with initialization, where a central server initializes a global model and distributes it to participating clients. Clients then perform local training on their private datasets, generating local updates such as gradients or model weights. These updates are subsequently transmitted to the central server through secure communication channels. The server performs aggregation by combining client updates into a unified model representation using methods such as Federated Averaging. The global model update stage follows, where the aggregated updates refine the global model before it is redistributed to clients for further training.

This cycle of local training, transmission, aggregation, and updating is repeated iteratively over multiple communication rounds until the model converges.

A. NOISE INTRODUCTION IN FEDERATED LEARNING

To address privacy concerns in FL, noise is strategically introduced at specific stages of the FL pipeline by leveraging the principles of Differential Privacy (DP). By injecting calibrated noise into client updates prior to aggregation, the system effectively obfuscates sensitive information such as identifiable patterns or membership signals, thereby reducing the risk of inference attacks [9]. At the same time, careful tuning of the noise parameters ensures that the model retains its predictive capacity, striking a balance between privacy and utility during the FL process [8], [26].

The categorization of noise injection into client-side, server-side, and bidirectional stages reflects insights from recent literature and the authors' interpretation aimed at systematically analyzing privacy–utility trade-offs in FL. Each position offers unique trade-offs in terms of privacy and utility.

1) Client-Side Noise Addition

This stage requires noise injection after local training and before transmitting updates to the server. In this approach, noise is added to local updates (gradients or weights) to protect sensitive client information [29]. The objective is to ensure that client data privacy remains intact even if the server is compromised. This aligns with the principle of local DP [26].

2) Server-Side Noise Addition

Noise is added during the aggregation of updates from multiple clients. For example, Gaussian noise can be applied to aggregated updates to prevent reconstruction of individual client contributions [30]. The objective of this approach is to maintain a balance between privacy and global model consistency, particularly when combined with secure aggregation methods [22].

3) Bidirectional Noise Introduction

This approach combines noise injection at both client and server levels, thereby providing layered privacy guarantees and increased robustness against inference and reconstruction attacks [31].

Real-world applications of FL include privacy-preserving healthcare collaborations, where sensitive patient data remain on local devices while still contributing to global models for tasks such as disease prediction, medical imaging, or diagnosis [32], [33]. Open-source libraries, such as TensorFlow Federated and PySyft, have emerged as key enablers for implementing these frameworks, offering built-in support for noise injection and secure aggregation.

Building on these foundations, the next section introduces our proposed framework, which systematically ingests noise at different stages of the FL lifecycle to evaluate its impact on privacy and utility within healthcare scenarios.

VI. PROPOSED FRAMEWORK

A. FEDERATED LEARNING WITH PRIVACY-PRESERVING NOISE INJECTION

This study extends traditional FL by integrating a privacy-preserving noise-injection mechanism to balance the trade-off between data utility and data privacy. As shown in Fig. 3, the proposed framework comprises the following key components: synthetic data generation, local model training, privacy-aware noise injection, secure model aggregation, performance evaluation, Frobenius distance calculation, and visualization of noise effects.

Components and Working of Fig. 3.

At the start of round t, the server broadcasts $W^{(t)}$ to a random subset of clients. Each client performs one local epoch and, in our implementation, applies client-side input of features (Age, Billing) prior to training, producing privatized updates. Clients return their (noisy) updates; once at least k updates are received (the *threshold* check), the server aggregates them with Krum to mitigate outliers and obtains $W^{(t+1)}$. The server then evaluates utility in terms of mean square error (MSE), mean absolute error (MAE), logs a privacy–utility monitor (PUL), and measures the Frobenius distance $\|W^{(t+1)} - W^{(t)}\|_F$ as a stability/perturbation indicator. Rounds repeat until the epoch budget is exhausted or a convergence criterion is met. While the figure shows both client-side and server-side noise injection points, only the client-side feature-level LDP path is active in our experiments; server-side perturbation is included to situate our design within the broader literature.

A synthetic dataset was created to emulate realistic healthcare scenarios using Python-based simulation tools, such as NumPy [34] for numerical features and the Faker library [35] for categorical and textual attributes. The generated features included name, age, sex, hospital and billing amount. This approach enabled controlled experimentation without compromising real patient data.

Clients then performed local training on their private synthetic datasets, which included sensitive attributes such as age, gender, diagnosis, and billing information. Only model updates were transmitted to the central server, reflecting the standard FL paradigm [2].

The next step is privacy-aware noise injection. In this work, noise is systematically injected on the *client side at the input feature level*, perturbing attributes such as age and billing before local training. This design ensures that privacy protection is applied at the data level and that even an honest-but-curious server observing client updates cannot infer raw data [26]. The framework quantified the privacy–utility trade-off by measuring stronger privacy guarantees (e.g., larger noise scales), which inversely impacted utility. Utility is measured using MSE and MAE, supporting the identification of noise parameters that achieve both strong privacy guarantees and acceptable accuracy in real-world healthcare applications. Although gradient-level noise and server-side noise injection have been proposed in the literature [30], these were not implemented in our experiments and are left as future work.

Secure model aggregation is performed using the Krum algorithm, a method that selects updates deviating the least from others to filter outliers [13].

Performance evaluation (utility) is conducted using regression metrics such as MSE and MAE. A custom Privacy–Utility Loss (PUL) function is also introduced to capture the trade-offs. Frobenius Distance is employed to quantify similarity between weight matrices across training rounds, thereby assessing the impact of noise on model convergence and stability [5]. The analysis thereof was used to assess how noise injection perturbs model updates, providing insight into both the stability and privacy effects of each noise mechanism. Finally, visualization of noise effects is performed to inspect data distributions and analyze how each noise type influences input variance. These visualizations provided insights into learning dynamics under privacy constraints and guided the subsequent analysis of privacy–utility trade-offs.

B. NOISE TRANSFORMATION AND UTILITY PRIVACY TRADE-OFF IN FL

Noise transformation in the implementation is achieved by introducing random perturbations drawn from well-defined statistical distributions to satisfy differential privacy requirements [8], [36]–[38]. The study considers five distributions:

Laplace Noise:

$$f(x) = \frac{1}{2b} e^{-\frac{|x|}{b}} \tag{1}$$

where b is the scale parameter [39]. This distribution is symmetric and heavy-tailed, making it suitable for privacy-preserving mechanisms.

Gaussian Noise:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \tag{2}$$

where σ is the standard deviation [40]. Gaussian noise provides smooth perturbations with predictable variance.

Exponential Noise:

$$f(x) = \lambda e^{-\lambda x} \tag{3}$$

where λ is the rate parameter. This asymmetric distribution is effective in scenarios requiring skewed perturbations [41].

Uniform Noise: Uniformly distributed in the range $[a, b]$, offering predictable transformations suitable for bounded perturbations [42].

Poisson Noise:

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{4}$$

a discrete distribution appropriate for count-based applications [43].

Each noise type uniquely influences data transformation, providing flexibility in balancing privacy and utility.

C. NOISE APPLICATION AND DATA PERTURBATION

The perturbation process can be expressed as:

$$\tilde{X} = X + n, \tag{5}$$

where X denotes the original data, $\tilde{X}$ the perturbed data, and $n \sim \mathcal{D}(\theta)$ is noise drawn from a chosen distribution $\mathcal{D}$ with parameters θ. The magnitude of θ determines the privacy–utility trade-off: larger values increase randomness and enhance privacy, while smaller values preserve more of the original signal (enhance utility).

VII. NOISE CALIBRATION AND TRADE-OFF ANALYSIS

THE following section evaluates different types of noise: Gaussian, Laplace, Exponential, Uniform, Poisson, and no-noise. The implementation provides a controlled testbed to analyze the impact of different noise types on model convergence. The experiments compared Laplace, Gaussian, Poisson, Exponential, and Uniform noise to determine their effectiveness under varying FL scenarios. These span various noise scales to determine the optimal noise scale that balances privacy (represented by the PUL) and utility (represented by the MSE). Table 7 presents the key results.

The detailed per-noise discussion (Gaussian, Laplace, Exponential, Uniform, Poisson) is provided in Appendix B.

A. DISCUSSION

As expected, the results shows that the absence of noise provides the lowest MSE (0.3209), and PUL (0.3900) provided minimal privacy protection. This observation emphasizes the necessity of adding noise to preserve privacy without losing the utility of data in privacy-critical scenarios.

Based on the empirical analysis, we recommend that all five types of noise use a scale of $(5, 3000)$ for future experiments. The experiments in the following sections use the recommended scale to maintain a balance between privacy and utility. This ensures reliable, repeatable, and reproducible results.

VIII. EXPERIMENTAL SETUP

THE experimental evaluation was conducted in a controlled FL environment designed to emulate a realistic healthcare setting while ensuring reproducibility. A total of six simulated clients participated in each experiment, contributing local updates to a central server across five communication rounds. In each round, three clients were randomly selected to participate, provided that the minimum aggregation threshold of three clients was satisfied. Local training was performed with a batch size of sixteen, and reproducibility was maintained by fixing the random seeds of NumPy, Python's `random` module, and TensorFlow.

The dataset comprised synthetic healthcare records generated using the `Faker` library combined with NumPy for numerical attributes. A total of 1000 records were distributed across the six clients, while a separate validation set of 500 records was reserved for the server. Each record included attributes such as name, age, gender, hospital, and billing amount. For model training, only the age and billing amount fields were used as input features, forming a two-dimensional feature space. Regression labels were synthetically generated as independent samples from a uniform distribution, $y \sim \mathcal{U}[0, 1)$.

Extended details on noise parameterizations, model definition, metric formulas, and visualization outputs are provided in Appendix C; the EDA figures and descriptive statistics tables are in the same Appendix (Fig. 10–14, Table 9).

1) Rigor Protocols for Gaussian LDP (for R3/R5)

Unless otherwise stated, we use Gaussian Local Differential Privacy (LDP) on normalized features (Age, Billing) with bounds Age ≤ 100, Billing $\leq 20{,}000$, target $\delta = 10^{-5}$, and analytic calibration of the Gaussian mechanism to a chosen ϵ. After LDP, no additional feature perturbation is applied during training (i.e., `-noise no_noise`). To strengthen experimental rigor we report:

- Variance across seeds: five independent runs at $\epsilon = 4$ with seeds $\{1, 2, 3, 4, 5\}$ (Table 11).
- Sensitivity to privacy level: an ϵ–sweep $\{2, 3, 4, 6, 8\}$; we record the calibrated σ and final metrics (Table 12).
- Non-IID stress: an age-skewed dataset created by sorting by Age to induce client-side imbalance; training proceeds identically (Table 13).

Attacker evaluation follows Yeom's loss-threshold Membership Inference Attack (MIA) [44]: we sweep the threshold τ and report the point of maximal advantage $\mathrm{Adv} = \mathrm{TPR} - \mathrm{FPR}$ together with True Positive Rate (TPR)/False Positive Rate (FPR).

IX. EMPIRICAL ANALYSIS OF DIFFERENT TYPES OF NOISE

THIS section describes the framework proposed in Section VI for evaluating the effectiveness of different types of noise in FL. The following are the key stages (initialization, local training, aggregation, and global update) of the FL pipeline used in the empirical analysis: Initiation ensures that a global model is initialized on the server and is distributed to the participants. The participants are then trained on their local private datasets to introduce noise to the updates. To enhance robustness against statistical outliers, global model aggregation was performed using the Krum algorithm, which filters extreme updates to stabilize model convergence under noisy conditions. No explicit adversarial or Byzantine scenarios were simulated, as the focus of this study is privacy preservation rather than attack resilience. Finally, the aggregated updates refine the global model and repeat multiple rounds until convergence is achieved.

A. MATHEMATICAL FORMULATION OF NOISE INTRODUCTION

For client i, the local updates u_i are perturbed by noise n_i such that:

$$\tilde{u}_i = u_i + n_i \tag{6}$$

where n_i is sampled from a chosen distribution (e.g., a Gaussian distribution with mean $\mu = 0$ and variance σ^2). The perturbed updates $\tilde{u}_i$ are sent to the server to ensuring obfuscation [8].

During aggregation at the server, perturbed updates are aggregated to form a global update:

$$G = \frac{1}{N}\sum_{i=1}^{N}\tilde{u}_i \tag{7}$$

where N is the total number of participating clients.

Significance in FL:

For Privacy, noise masks the contribution of individual client updates and mitigates privacy risks [29]. With respect to Utility, adjusting the noise scale (for example, the standard deviation in Gaussian noise) controls the inverse privacy–utility relationship; increasing the standard deviation improves privacy by better masking individual updates. Conversely, it also lowers utility, as shown by metrics such as MSE [30].

B. HOW VISUALIZATIONS INFORM PRIVACY AND UTILITY ANALYSIS

Scatter plots of the original and noisy data provided critical insights into privacy and utility. When noisy data points remain close to their original positions, utility is preserved because the model can still infer meaningful patterns. Furthermore, a large dispersion of noisy points indicates greater deviation from the original data, leading to reduced utility (higher MSE).

With respect to privacy, a greater dispersion of noisy points increases privacy by making it harder for adversaries to reverse engineer original data. However, clustering or predictable shifts (e.g., in exponential noise) may leak certain patterns that pose privacy risks.

C. BASELINE EXPERIMENT: NO NOISE

The baseline experiment presented in Fig. 4 evaluates the impact of the noise mechanisms. In this setup, no noise is applied to the data. This results in the maximum utility and no privacy enhancement.

The baseline data remains unchanged, the equation is:

$$\tilde{X} = X \tag{8}$$

1) Findings from the Baseline Experiment

The visualization of the baseline experiment (shown in Fig. 4) highlights the following:

In terms of Utility, the dense clustering of dots around true data points reflects the absence of distortions. This reflects the maximum accuracy of the predictive tasks.

In terms of Privacy, in the absence of noise injection, local model updates can retain patterns that are directly correlated with the individual data points. These patterns can be exploited via inference attacks. For example, membership inference and gradient inversion have been used to extract sensitive information about patients [17], [45].

TABLE 1: Sampled Records from Baseline Experiment

Original Values		No-Noise Values	
Age	Billing	Age	Billing
60	11647	60	11647
37	11735	37	11735
94	5854	94	5854
82	12543	82	12543
88	9168	88	9168

D. EXPONENTIAL NOISE

Exponential noise introduces asymmetry in perturbed data. Often, data are skewed in one direction. This characteristic (biased skewed distribution) is ideal for scenarios in which the directional sensitivity is less critical.

$$\tilde{x} = x + \text{Exponential}(\lambda) \tag{9}$$

where λ is the rate parameter controlling the noise scale.

To quantify the utility loss, the MSE is calculated as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \tilde{x}_i)^2 \tag{10}$$

The Frobenius distance d_F between the original and noisy datasets is given by equation 16 in section C

1) Findings from the Exponential Noise Experiment

With respect to Utility, Fig. 5 shows skewed perturbation biases in the data. This initially reduces the model accuracy through system over- or underestimation. In earlier epochs, the skewed bias translated to slightly higher prediction errors (as observed by the increased MSE and MAE discussed in Section X) compared with the baseline. However, during training, the model slightly adapted to skewed data distributions. The positively skewed distribution was reflected by a slight decrease in PUL by the final epoch. Overall, Exponential noise incurs a moderate utility loss, more than symmetric (baseline, no noise) noise, which has no net bias but is not large enough to prevent the model from learning effectively.

Regarding Privacy, the asymmetry of the exponential noise implies that each original value is consistently offset by a positive value. This makes it difficult to infer the true values. An adversary can only state that the original value is lower than the noisy value but not how much lower. The divergence between the noisy trained model and baseline model grows across epochs. The Frobenius distance difference was significant (approximately 3.9 at epoch 5), confirming that the noise significantly perturbs the model parameters. This substantial perturbation provides privacy by decoupling the model's learned parameters from the precise original data values.

TABLE 2: Five sample records: original vs. noisy Age and Billing values under Exponential noise. Exponential noise tends to add a positive offset (one-sided), evident from the generally increased Billing values and variable shifts in Age.

Original Values		Exponential Noise Values	
Age	**Billing**	**Age**	**Billing**
60	11647	56.02	11872.71
37	11735	35.90	12270.66
94	5854	99.32	8529.32
82	12543	83.59	10404.72
88	9168	91.21	8184.65

E. GAUSSIAN NOISE

The Gaussian noise introduces symmetric perturbations and aligns with the central limit theorem. This implies that it can be widely used in privacy-preserving FL. The perturbed data are defined as:

$$\tilde{x} = x + \mathcal{N}(0, \sigma^2) \tag{11}$$

where $\mathcal{N}(0, \sigma^2)$ is a Gaussian distribution with a mean of zero and variance σ^2.

1) Findings from the Gaussian Noise Experiment

For Utility, Gaussian noise adds zero-mean fluctuations to the data (Fig. 6). Therefore, positive and negative perturbations tend to cancel out across the datasets. Visually, the noisy data points remained distributed around the true values without a systematic bias, resulting in a moderate spread around each original point. Consequently, the model experiences minimal utility degradation: early round prediction errors (MSE and MAE) with Gaussian noise data are almost identical to the baseline, and by the final epoch, the accuracy of the model is essentially unchanged. The PUL metric remained nearly constant throughout training and improved slightly by epoch 5, indicating that the model could absorb Gaussian noise without significant long-term accuracy loss.

With respect to Privacy, symmetric random perturbations from Gaussian noise substantially enhance the privacy. This is mainly because the observed noisy value can be either higher or lower than the actual value with equal likelihood. This unpredictability makes it difficult for an attacker to infer the original data from the output after noise has been injected. The scatterplots confirm a broad but centered dispersion of points, meaning that sensitive attributes, such as Age and Billing, are obscured by moderate random noise. The impact on the model is evident in the growing difference between the noisy and noise-free models over time (final Frobenius Distance difference ≈ 3.9), which indicates that Gaussian noise injection significantly and meaningfully alters the learned model parameters to protect privacy, while maintaining model fidelity.

The data in Table 3 shows that Gaussian noise is symmetric, leading to some values increasing and others decreasing (e.g., Billing amounts show both upward and downward noisy deviations).

TABLE 3: Five sample records: original vs. noisy Age and Billing values under Gaussian noise.

Original Values		Gaussian Noise Values	
Age	**Billing**	**Age**	**Billing**
60	11647	60.89	14889.34
37	11735	30.32	14236.77
94	5854	95.90	7231.54
82	12543	85.05	12332.50
88	9168	90.80	4185.12

F. LAPLACE NOISE

Laplace noise is generated using a Laplace distribution and is widely used in DP. The noise is defined as:

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right) \tag{12}$$

where μ is the mean and b is the scale parameter.

1) Findings from the Laplace Noise Experiment

With respect to Utility; Laplace noise, owing to its heavier tails, occasionally produces larger perturbations than Gaussian noise does. In the scatterplots (Figure 7), the noisy data points were still centered around the true values but with more frequent outliers; some points were significantly displaced. These larger deviations (for example, a Billing value shifting by several thousand) can cause a slight increase in the prediction error relative to the Gaussian noise. Despite this, the model continued to learn from the data; over the training epochs, the performance with Laplace-noised data approached that of the baseline model. The PUL remained steady across epochs, dropping only by the final epoch, indicating that the model compensated for much of the noise-induced error by convergence.

With respect to Privacy; Laplace noise provides strong privacy protection. This is expected from the mechanism that is widely used in DP. Noise is symmetric about zero. Therefore, an attacker cannot easily discern whether a noisy value is above or below the true value. Moreover, the heavier tail of the Laplace distribution means that there is a higher chance of large noise values, which effectively conceal the original data points (some original values are masked by large positive or negative jumps). The model parameters diverged from those of the noise-free case, similar to the Gaussian scenario (final Frobenius Distance difference ≈ 3.9), reflecting a significant noise injection. Overall, Laplace noise achieves high privacy guarantees at the cost of a moderately higher utility loss than that of Gaussian noise.

TABLE 4: Five sample records: original vs. noisy Age and Billing values under Laplace noise.

Original Values		Laplace Noise Values	
Age	**Billing**	**Age**	**Billing**
60	11647	55.03	12793.27
37	11735	37.44	13191.22
94	5854	100.85	9449.87
82	12543	85.12	10459.57
88	9168	92.75	9103.87

Table 4 shows that Laplace noise (with its heavy-tailed distribution) causes some values to change substantially (see the large Billing value shifts), though many remain near the original.

G. POISSON NOISE

Poisson noise is ideal for counting data and introduces variability that is proportional to the rate parameter:

$$P(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} \tag{13}$$

where k is the count and λ is the rate parameter.

1) Findings from the Poisson Noise Experiment

In terms of Utility; Poisson noise introduces small integer changes to the data proportional to rate λ. In our setting, these perturbations are minor for both the Age and Billing fields (Figure 9); therefore, the noisy data appear almost identical to the original distribution. For instance, an age of 37 could become 36, and a billing amount of 11647 could become 11679; such slight differences had a negligible effect on the model predictions. Accordingly, the utility of the model was essentially preserved; the MSE and MAE remained almost unchanged when Poisson noise was added, and the PUL was very low from the first epoch. By the final epoch, the model trained on Poisson-noised data was on par with the baseline model, indicating virtually no loss of accuracy.

The Privacy benefit of Poisson noise is modest given the small magnitude of noise. It is well suited for discrete count data scenarios — an adversary can no longer be certain of an exact count. However, in our case (with relatively small noise values), they can still estimate the original data fairly closely. However, even small per-update perturbations accumulate uncertainty; over multiple training rounds, the difference between the noisy data model and the original model becomes noticeable. By epoch 5, the Frobenius Distance gap reached approximately 3.9, which is comparable to other noise types, indicating that the repeated application of Poisson noise over the training process introduces a meaningful overall change. In summary, Poisson noise offers excellent utility retention with limited but nonzero improvement in privacy.

TABLE 5: Five sample records: original vs. noisy Age and Billing values under Poisson noise.

Original Values		Poisson Noise Values	
Age	Billing	Age	Billing
60	11647	63.00	11679.00
37	11735	36.00	11772.00
94	5854	94.00	5824.00
82	12543	87.00	12593.00
88	9168	86.00	9188.00

Table 5 shows that Poisson noise changes each value by only a small amount (usually a few units), preserving the data's general privacy (note that Ages and Billing amounts remain very close to their original values).

H. UNIFORM NOISE

Uniform noise evenly perturbs data points within a fixed range:

$$f(x) = \frac{1}{b-a}, \quad a \le x \le b \tag{14}$$

1) Findings from the Uniform Noise Experiment

For utility purposes, uniform noise perturbs data uniformly at random within a specified range. This resulted in a broad dispersion of perturbed data points around the originals (Figure 8). Visually, noisy values are spread out more widely than with any other mechanism, because every value in the range is equally likely to occur. Some points exhibited large positive or negative shifts. Such unstructured, potentially large deviations can noticeably degrade data utility. Indeed, in the initial epochs, the model trained on uniformly noisy data showed higher error rates than those trained with other noises. However, the federated model partially overcomes this noise. By the final epoch, the MSE and MAE of the uniform-noise model nearly matched the baseline, indicating that the model adapted and recovered some lost utility. Correspondingly, PUL, which was the highest initially for uniform noise, decreased by the end of the training, although uniform noise still generally imposed the highest utility cost among the noise types.

In terms of privacy, the strength of the uniform noise lies in its maximum uncertainty. Because the noise values were uniformly distributed in a range, an observed noisy value could have originated from any original value within that range. This makes it extremely difficult for an attacker to infer the true value (the original value can be anywhere in the interval around a noisy observation). In our experiments, the uniform noise range was sufficiently large to cause some outputs to deviate substantially from their true values (e.g., a billing amount that changed by more than a thousand). The effect on the model is a significant divergence from the baseline training (final Frobenius Distance difference ≈ 3.9), which is similar in magnitude to other noise methods. In exchange for this high level of privacy (the original data are well hidden), uniform noise exacts the heaviest toll on utility when the noise range is wide.

TABLE 6: Five sample records: original vs. noisy Age and Billing values under Uniform noise.

Original Values		Uniform Noise Values	
Age	Billing	Age	Billing
60	11647	56.85	12599.70
37	11735	37.42	12888.66
94	5854	97.73	7949.17
82	12543	84.32	11041.01
88	9168	91.07	9104.55

The noise is evenly distributed across a range, so some values are only slightly changed (e.g., Age 37 to 37.42) while others vary more dramatically (e.g., Billing 5854 to 7949.17), illustrating the comparatively high distortion from uniform noise.

I. KEY OBSERVATIONS

The combined visual and quantitative analyses highlight the distinct privacy-utility trade-offs of each noise mechanism. For the baseline, where there is no noise, the results show maximum utility (i.e., the model sees the exact data), but there is no privacy protection. Gaussian Noise achieves the best overall privacy-utility balance, introducing relatively small symmetric distortions that barely affect the model accuracy while still obscuring individual data points. The third observation is that Laplace Noise provides strong privacy guarantees through the possibility of larger perturbations (i.e., heavy-tailed noise) at the cost of a moderate increase in the prediction error (i.e., utility loss) compared to Gaussian Noise. Exponential Noise offers directional privacy by constantly shifting values in one direction. However, this induces bias in the data, which can affect utility and prediction fairness. The Poisson Noise is ideal for discrete count data. It adds minimal noise that almost achieves the maximum utility. However, the privacy gain was limited, because the perturbed values remained very close to the original values. Finally, Uniform Noise delivers high privacy guarantees by uniformly spreading values over a wide range, making it difficult to guess the original data. However, it achieves the greatest utility loss, owing to its potential for large distortions.

In summary, although all noise types eventually led to a similar magnitude of model parameter perturbation after training (comparable final Frobenius Distances), their immediate impacts on the model accuracy varied. Gaussian and Poisson noise had a negligible impact on performance, Laplace and Exponential noise imposed slightly higher costs, and Uniform noise showed the most significant initial utility degradation. These findings underscore the importance of selecting a noise mechanism that is appropriate for data and privacy requirements. For example, Gaussian or Poisson noise may be preferable when model accuracy is paramount, whereas Laplace or Uniform noise may be chosen when stronger privacy guarantees are needed, despite the higher utility trade-off.

X. COMPREHENSIVE EMPIRICAL ANALYSIS

THIS section presents the empirical findings derived from experiments involving five types of noise: Exponential, Gaussian, Laplace, Poisson, and Uniform, and a no-noise baseline. The results were obtained from simulated FL scenarios designed to evaluate the trade-off between privacy and utility. The section begins by outlining the metrics selected for analysis and explaining their relevance and its application in an experimental context. This is followed by a comprehensive evaluation of how each noise mechanism affects FL performance through the metrics outlined in the next subsection.

A. METRICS OVERVIEW

The experiments used the MSE metric, which quantifies utility loss by measuring the average squared difference between the predicted and true values. A lower MSE indicates a higher utility, mainly because the predictions are closer to the ground truth. The second metric is the MAE. This metric reflects the absolute average error between the predictions and actual values and serves as an intuitive measure of data utility. The third metric is the Mean Bias Deviation (MBD), which measures the systematic bias introduced by noise. Positive or negative values reflect whether the noise consistently shifts predictions in one direction or another. However, for the purposes of this study, we did not use the MBD metric.

The next metric is the Privacy-Utility Loss (PUL), which combines the MSE and a privacy penalty (ϵ) proportional to the noise scale (α). This is defined as:

$$\text{PUL} = \text{MSE} + \epsilon \cdot \alpha \tag{15}$$

This metric allows for a direct comparison of the trade-off between privacy and utility.

The last metric is the Frobenius Distance, which measures the divergence (d_F) between global model weights (w_i) across epochs. This is defined as shown by equation 16 in C. Larger distances indicate greater model perturbation (i.e., deliberate alterations introduced to data or model parameters, typically by adding noise to preserve privacy) owing to the noise [46]. To better understand how such perturbations affect data utility, the following section delves into how metrics such as MSE and MAE reflect utility in privacy-preserving FL scenarios.

Regarding the interpretation of MSE or MAE against Utility, a lower MSE or MAE signifies that the predictions are closer to the actual values, which reflects a higher utility. This relationship holds because predictions with smaller deviations from the ground truth maintain the fidelity of the original data. This is critical for enhancing the utility. In machine-learning tasks, higher utility is synonymous with lower error rates in predictions, which are directly measured by MSE or MAE.

In contrast to utility, which is measured using MSE or MAE, privacy is inferred using the Frobenius Distance. The Frobenius distance measures the changes in the global model weights across epochs. Higher distances reflect more significant changes, usually owing to noise perturbation, which enhances privacy. This is because larger perturbations make it more difficult for an adversary to reverse-engineer the original data from the model weights. Furthermore, the divergence in weights ensures that individual contributions to the model are obscured, aligning with the privacy goals.

B. EXPERIMENT FINDINGS BY NOISE TYPE

The baseline experiment with No Noise served as a reference point for evaluating model performance without any noise-induced perturbations. As shown in Table 14, this configuration achieved a final MSE of 0.321 and MAE of 0.482. These metrics indicate high predictive accuracy and essentially no systematic bias in the model's predictions. PUL remained extremely low (approximately 0.371 in the final epoch). The changes in Frobenius distance were minimal in each round.

This demonstrates stable and consistent model updates in the absence of noise.

Exponential Noise introduces one-sided (positive) perturbations into the data, which can consistently shift values. In our experiments, adding exponential noise resulted in a final MSE of 0.323 and MAE of 0.485. This is only a marginal increase compared with the no-noise baseline. The PUL for this setting is 0.373 during the last epoch. The Frobenius Distance of the model update reached approximately 3.88 by epoch 5. This indicates that the cumulative effect on the model parameters is comparable to that of the other noise types. These results suggest that exponential noise had a noticeable but limited impact on utility and did not destabilize training. Furthermore, it supports a relatively balanced trade-off between privacy and accuracy in the FL process.

Experiments conducted on Gaussian Noise added symmetric zero-mean perturbations to the data, leading to minimal overall distortion. According to Table 15, this approach yielded an MSE of 0.325 and MAE of 0.487 in the final model. This error was nearly as low as the baseline error. The corresponding PUL is 0.373. This is one of the lowest noise mechanisms. The Frobenius distance for Gaussian Noise reached roughly 3.88 by the end of the training, the Frobenius distance for the no-noise scenario. These metrics confirm that Gaussian Noise effectively enhances utility, while introducing sufficient random perturbation to provide privacy guarantees. Its balanced nature makes it particularly suitable for privacy-preserving machine-learning settings, where high model fidelity must be maintained.

Owing to its heavily skewed distribution, Laplace Noise generates large but infrequent perturbations. In our results, this noise led to a final MSE of 0.321 and MAE of 0.483, essentially matching the baseline performance. The PUL stabilized at 0.373 by epoch 5, and the Frobenius distance reached approximately 3.88, which is similar to that in the Gaussian case. Although the heavily skewed nature of Laplace Noise may introduce more variability during training (through occasional outlier updates), these figures demonstrate that strong privacy guarantees can be achieved without significant long-term utility degradation. Thus, Laplace Noise can be argued to offer robust privacy (owing to its potential for larger noise values), with only a negligible sacrifice in utility, albeit with a slightly less predictable training trajectory.

Experiments with Poisson Noise, which are particularly suited for discrete or count data, show a moderate level of perturbation for each value. The experimental results in Table 18 show a final MSE of 0.328 and MAE of 0.491 for this setting, making it the highest error among the tested types of noise. The PUL converged to 0.378, which is slightly higher than that of other noise distributions, such as Exponential and Laplace distributions. The Frobenius distance in the final epoch is approximately 3.88. These findings indicate that, while Poisson Noise is effective in masking individual counts (preserving privacy for integer-valued data), it introduces a small additional utility cost and marginally greater variability in model updates. In practice, Poisson Noise may cause minor shifts in model behavior, which must be accounted for when the maximum utility is critical.

Finally, Uniform Noise-applied perturbations are evenly drawn from a fixed range. This implies that all values within a certain interval are equally likely to be noise offset. This broad-spectrum perturbation has the potential to degrade the model performance more than other noise types, particularly in the early stages of training. However, by the final epoch, the model with Uniform Noise achieved an MSE of 0.323 and an MAE of 0.485, which are similar to the results obtained with Gaussian and Laplace Noise. Similarly, the PUL settled at 0.373, and the Frobenius Distance of the parameter changes reached approximately 3.88. This is comparable to the other noise scenarios. These results emphasize that although Uniform Noise is not biased towards small perturbations (increasing the risk of larger individual deviations), with sufficient training, it is possible to recover most of the lost utility. Nevertheless, Uniform Noise should be used with careful calibration, as an improperly chosen noise range can significantly compromise the model consistency and accuracy.

C. TABULATED EMPIRICAL ANALYSIS OF DIFFERENT TYPES OF NOISE

The key results are summarized in Tables 8 to 19, highlighting the model accuracy (MSE, key metric for utility), PUL, and magnitude of the model update divergence (Frobenius Distance) for each noise setting in the final training epoch.

D. PRIVACY EVIDENCE: FORMAL ACCOUNTING AND ATTACKER EVALUATION

DP–SGD configuration: Clipping norm $C{=}1.0$, Gaussian noise multiplier $\sigma{=}1.2$, batch size $B{=}16$, 6 total participants with $K{=}3$ selected per round, 5 epochs, and target $\delta{=}10^{-5}$. The accountant composes privacy loss over optimizer steps (Rényi orders scanned and converted to (ϵ, δ)); we report the worst ϵ observed across rounds.

LDP calibration: Features (Age, Billing) are clipped to known bounds (Age ≤ 100, Billing $\leq 20{,}000$), normalized to $[0, 1]$, and privatized with a calibrated Gaussian mechanism to a target $(\epsilon{=}4, \delta{=}10^{-5})$; the run records the calibrated σ and bounds in `ldp_meta.json` for provenance.

Attacker: We implement the loss-threshold membership inference attack of Yeom *et al.* [44]: for a threshold τ, predict "member" if loss $\leq \tau$. We sweep τ and report the point of maximal advantage $\mathrm{Adv} = \mathrm{TPR} - \mathrm{FPR}$ (the achieving τ is shown in-text).

Interpretation. As per Table 10, DP–SGD attains a formal guarantee of $\epsilon{\approx}4.22$ at $\delta{=}10^{-5}$ with validation error comparable to the non-private baseline. Gaussian LDP achieves the target $\epsilon{=}4$ (local guarantee) with similar utility in this synthetic task. Yeom MIA advantages are small in both cases (DP–SGD: $\mathrm{Adv}{=}0.020$ at $\tau{\approx}0.318$; LDP: $\mathrm{Adv}{=}0.033$ at $\tau{\approx}0.288$), providing an e1mpirical sanity check alongside formal accounting.

1) Pertubations (no formal DP).

Noise perturbations such as (for example: baseline noise) with ad-hoc scales (e.g., Age ± 5, Billing ± 3000) are *utility probes only* and do not constitute DP unless noise is *calibrated from sensitivity to a target* (ϵ, δ) and composed with an accountant. We therefore report these perturbations as proxy observations and reserve privacy claims for the formal DP rows in table 10.

E. GAUSSIAN LDP: VARIANCE, SENSITIVITY, AND NON-IID ROBUSTNESS

Variance across seeds (Table 11). At $\epsilon = 4$ the final-epoch utility is stable: MSE $= 0.3307 \pm 0.0160$, MAE $= 0.4981 \pm 0.0145$, PU $= 0.3860 \pm 0.0159$, and Frobenius 3.174 ± 0.686. MIA advantage remains small (0.079 ± 0.0367), indicating limited membership signal under the formal LDP guarantee.

Sensitivity to ϵ (Table 12). Tightening privacy (smaller ϵ, larger σ) mildly increases error (for example, $\epsilon = 2$: MSE 0.3498), whereas relaxing privacy preserves or improves utility ($\epsilon = 8$: MSE 0.3216). Across the sweep, MIA advantages stay low (0.025–0.137), providing an empirical sanity check that aligns with the formal (ϵ, δ) accounting.

Non-IID stress (Table 13). Under age-skewed participation, utility degrades modestly (MSE 0.3387, MAE 0.5079) while MIA advantage remains small (0.080). This suggests the mechanism retains privacy–utility balance under benign distribution shift.

These results substantiate that *Gaussian LDP* delivers reproducible utility with bounded privacy leakage, scales predictably with ϵ, and is resilient to non-IID skew.

F. KEY OBSERVATIONS

The results indicate that the baseline (i.e., no noise) yields maximal data utility, but zero privacy protection. While ideal in terms of performance, such an approach is unsuitable for sensitive domains (e.g., healthcare or finance), where strict data confidentiality is mandated by regulations such as The Health Insurance Portability and Accountability Act (HIPAA), General Data Protection Regulation (GDPR), and Protection of Personal Information Act (POPIA). The baseline scenario underscores the necessity of incorporating privacy mechanisms into FL despite the slight performance trade-offs that they entail.

The results also demonstrate that Gaussian noise provides an excellent trade-off between privacy and utility. At the end of the training, this indicates that it had no direct impact on the accuracy of the data. This makes Gaussian noise particularly suitable for applications such as healthcare analytics and wearable sensor data, where moderate noise can protect personal information without compromising the quality of insights derived from the original data.

Laplace noise offers strong privacy guarantees owing to its occasional large perturbations; however, it still converges to high accuracy with only a minor utility cost. This type of noise is advantageous in high-risk scenarios (such as whistleblower platforms or military communication systems), where maximizing privacy and resistance to inference attacks is crucial, even if it implies slightly more variability during training.

Finally, Uniform noise introduced the most unpredictable perturbations. This highlights the need for careful calibration of the magnitude of the noise. Although our final results show that the model can recover much of its performance, uniform noise can cause a significant utility loss if misconfigured. It may be viable in applications where individual precision is less important than anonymity (e.g., large-scale social media trend analysis). However, it must be used cautiously to avoid overwhelming the model with noise.

These findings confirm that the choice of noise distribution has a measurable influence on the federated data utility and privacy. They also illustrated that given sufficient training and appropriate noise scaling, even aggressive noise mechanisms can be balanced to achieve acceptable performance. This paves the way for a deeper exploration of noise-calibration strategies and their implications for real-world privacy-preserving FL deployment.

XI. DISCUSSION

THIS study aims to investigate how different noise mechanisms influence the delicate trade-off between privacy protection and data utility in FL. Privacy-preserving techniques in FL play a critical role in ensuring data confidentiality while maintaining model performance, particularly in sensitive application domains such as healthcare, where both dimensions are non-negotiable. This study systematically evaluates the effects of different noise-injection mechanisms—Gaussian, Laplace, Poisson, Exponential, and Uniform—on data utility and privacy. The results highlight significant trade-offs, reinforcing the need for strategic noise selection based on cybersecurity application-specific requirements.

A. IMPLICATIONS OF FINDINGS

The results demonstrate that Gaussian and Laplace noises provide the best balance between privacy and utility, making them strong candidates for practical deployment in healthcare information systems. Gaussian noise, which is known for its widespread use in DP, showed a stable model performance with minimal accuracy degradation. Laplace noise, while introducing a slightly higher distortion, provided stronger privacy guarantees owing to its heavier tails, making it suitable for adversarial robust environments.

Despite maintaining moderate stability, Poisson noise exhibited increased Frobenius Distances across epochs, suggesting a greater variance in participant updates and potential model divergence over multiple training rounds. This variability could pose challenges for FL in highly dynamic environments where model accuracy is crucial, thus making it crucial to balance the trade-off between data privacy and utility.

Exponential noise produces the highest level of privacy protection but at the cost of substantial model degradation. Aggressive perturbation led to a sharp increase in both MSE

and MAE, making it less practical for applications in which predictive accuracy is a primary concern.

Uniform noise exhibited the most unpredictable performance with inconsistent utility scores and fluctuating Frobenius Distances. This is because a uniform distribution injects noise of equal magnitude across its entire range, occasionally producing large perturbations that destabilize model updates. As a result, Uniform noise fails to deliver a consistent privacy–utility trade-off and can undermine model reliability in mission-critical domains, such as healthcare.

From a cybersecurity implementation perspective, the selection of a noise mechanism must align with the broader threat model and the operational security posture of the deployment environment. Gaussian and Laplace noise are particularly suited for integration into existing privacy-preserving infrastructures, as they offer quantifiable guarantees under DP frameworks commonly referenced in compliance standards, such as GDPR/POPIA and HIPAA. Furthermore, their predictable impact on model performance enables easier integration into risk-management workflows and security audits. In contrast, the instability associated with Uniform and Exponential noise may complicate threat modelling and reliability assurance, especially in mission-critical sectors, such as healthcare or finance, where reproducibility, traceability, and auditability are essential for certification and trust.

B. COMPARISON WITH EXISTING LITERATURE

These findings align with those of previous research that emphasized Gaussian and Laplace noise as the most effective choices for privacy-preserving FL. Prior studies have highlighted that Gaussian noise maintains a smooth perturbation distribution, resulting in lower distortion in the model weights. Laplace noise, although slightly more disruptive, is commonly adopted because of its compliance with strong DP guarantees [8], [9], [26], [47].

However, the existing literature has often emphasized privacy in isolation, without critically assessing how these noise mechanisms impact learning stability, defense robustness, and thus, data utility. Our study extends this understanding by introducing Frobenius Distance as a quantitative metric to measure model-update divergence, reflecting the privacy–utility trade-off induced by different noise distributions. This allows for a more cybersecurity-relevant interpretation of the noise. Excessive perturbation, as observed with Exponential and Uniform noise, may degrade convergence and increase vulnerability to attack vectors that exploit unstable model behaviors, such as adaptive poisoning or gradient leakage [48], [49].

From a cybersecurity standpoint, this study suggests that Gaussian and Laplace noise offer not only formal privacy guarantees, but also practical resilience against inference-based threats without significantly compromising data utility. The strong privacy and moderate utility preservation shown in our evaluation underscores the importance of choosing noise schemes that align with both privacy standards and real-world security requirements, especially in sensitive domains such as healthcare.

This study bridges the gap between theoretical privacy and practical security implementations by directly linking privacy-preserving mechanisms to model behavior over training epochs, an area that remains underexplored in FL research.

XII. CONCLUSION

THE insights gained from this study contribute to the growing body of research on privacy–utility-aware FL. As FL adoption increases, refining privacy-preserving techniques is essential to ensure participant data remains private. By leveraging a strategic combination of noise mechanisms, such as Gaussian and Laplace, together with advanced security techniques such as Byzantine-resilient aggregation (e.g., Krum) and anomaly-based client selection, FL can move closer to practical, large-scale real-world deployment.This study contributes to advancing privacy–utility-aware FL frameworks by providing empirical evidence of the effectiveness of various noise models in balancing these critical trade-offs.

This study explored the impact of different noise distributions on privacy and utility in FL, with a particular focus on healthcare applications in which patient confidentiality must be preserved without compromising the diagnostic performance. By evaluating the Gaussian, Laplace, Poisson, Exponential, and Uniform noise models, we assessed their effects on model performance using key metrics such as MSE and MAE. Furthermore, we analyzed privacy–utility trade-offs using Frobenius Distance metrics across training epochs. To the best of our knowledge, prior research has not utilized the Frobenius Distance in conjunction with visualization techniques to assess the impact of noise on FL. This dual approach represents a novel contribution of our study, offering both quantitative and qualitative insights into the privacy–utility trade-off.

Our findings, based strictly on the reported metrics, indicate that Gaussian and Laplace noises provided the most balanced trade-offs between privacy and utility. Gaussian noise demonstrated stable performance with minimal impact on model accuracy, whereas Laplace noise enhanced privacy proxies with slightly higher distortion. Poisson noise exhibited moderate stability but introduced higher Frobenius Distances, suggesting greater model divergence. Although Exponential noise offered stronger perturbation, it resulted in significant performance degradation, reducing its applicability in utility-sensitive environments. By contrast, Uniform noise produced inconsistent results, leading to reliability concerns.

From a practical standpoint, these insights can guide system architects and cybersecurity teams in selecting appropriate noise mechanisms based on deployment context. For instance, healthcare platforms prioritizing diagnostic accuracy may favor Gaussian noise to retain utility, whereas applications requiring stronger privacy protection may opt

for Laplace noise. The impact is quantifiable: utility was measured through predictive error metrics (MSE and MAE), whereas privacy resilience was reflected in reduced gradient similarity and higher Frobenius Distances. By evaluating both, researchers and cybersecurity practitioners can make informed trade-offs that align with their requirements and operational goals when implementing privacy-preserving techniques in sensitive domains such as healthcare.

A. FUTURE RESEARCH DIRECTIONS

Based on these findings, future research avenues may include the development of an intelligent noise-scaling mechanism that dynamically adjusts privacy budgets based on model convergence rates and sensitivity analyses. Future research may also include exploratory combinations of different types of noise-based perturbations with cryptographic approaches to preserve privacy without an excessive loss of utility. Furthermore, future research may evaluate how different noise mechanisms occur under adversarial scenarios such as gradient inversion and model poisoning attacks. The last recommendation is for research that extends the current noise analysis beyond healthcare to other domains, such as finance, smart grids, and industrial IoT, where FL may be applicable. Future research should extend the evaluations to multi-institutional applications with non-IID (nonindependent and identically distributed) data partitions. Future work should investigate how combining noise injection with cryptographic techniques (e.g., homomorphic encryption and secure multi-party computation) impacts both privacy and utility.

DECLARATION

This paper was written entirely by the authors, and the opinions, ideas, research, analysis or conclusions are solely those of the authors and cannot be attributed to their affiliated institutions. However, Large Language Models (LLMs) such as ChatGPT, Overleaf AI Assist, and Gemini have been used to correct spelling and grammatical errors. The authors affirm that AI tools were used solely for linguistic refinement and to generate synthetic data for experimentation.

XIII. REFERENCES

[1] IBM Corporation and Ponemon Institute, "Cost of a data breach report 2024," Armonk, NY, USA, 2024, accessed: 2025-05-12. [Online]. Available: https://www.ibm.com/reports/data-breach

[2] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," Jan. 2023, arXiv:1602.05629 [cs]. [Online]. Available: http://arxiv.org/abs/1602.05629

[3] S. Bharati, M. R. H. Mondal, and P. Podder, "Federated learning: Applications, challenges and future directions," *Journal of Hybrid Intelligence Systems*, 2022. [Online]. Available: https://journals.sagepub.com/doi/abs/10.3233/HIS-220006

[4] "Advances and Open Problems in Federated Learning." [Online]. Available: https://ieeexplore.ieee.org/document/9464278/authors#authors

[5] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How To Backdoor Federated Learning," Aug. 2019, arXiv:1807.00459 [cs]. [Online]. Available: http://arxiv.org/abs/1807.00459

[6] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 691–706.

[7] B. Zhao, K. R. Mopuri, and H. Bilen, "Idlg: Improved deep leakage from gradients," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 0–1.

[8] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep Learning with Differential Privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, Oct. 2016, pp. 308–318, arXiv:1607.00133 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1607.00133

[9] C. Dwork and A. Roth, "The Algorithmic Foundations of Differential Privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, Aug. 2014, publisher: Now Publishers, Inc. [Online]. Available: https://www.nowpublishers.com/article/Details/TCS-042

[10] Y. Zhao and C. Wang, "Protecting privacy and enhancing utility: A novel approach for personalized trajectory data publishing using noisy prefix tree," *Computers & Security*, vol. 144, p. 103922, 2024.

[11] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," 2018, arXiv:1806.00582 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1806.00582

[12] D. Yin, Y. Chen, K. Ramchandran, and P. Bartlett, "Byzantine-robust distributed learning: Towards optimal statistical rates," in *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

[13] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, "Machine learning with adversaries: Byzantine tolerant gradient descent," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[14] H. Su, Y. Zhang, H. Zhang, X. Liang, X. Zhang, and Z. Xu, "Trustfed: A reliable federated learning framework with malicious-attack resistance," *arXiv preprint arXiv:2312.04597*, 2023. [Online]. Available: https://arxiv.org/abs/2312.04597

[15] H. Yu, C. Ma, M. Liu, T. Du, M. Ding, T. Xiang, S. Ji, and X. Liu, "G^2uardfl: Safeguarding federated learning against backdoor attacks through attributed client graph clustering," *arXiv preprint arXiv:2306.04984*, 2023. [Online]. Available: https://arxiv.org/abs/2306.04984

[16] M. Nasr, R. Shokri, and A. Houmansadr, "Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning," in *IEEE Symposium on Security and Privacy (SP)*, 2019.

[17] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.

[18] L. Zhu, Y. Wang, and P. Kairouz, "Differential privacy for federated learning: A survey of recent advances," *ACM Computing Surveys*, 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3582295

[19] V. Nikolaenko, S. Halevi, and B. Pinkas, "Secure multi-party computation for large-scale federated learning," *Proceedings of IEEE Transactions on Information Forensics and Security*, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9637845

[20] X. Zhao, B. Li, and W. Du, "Enhanced secure multi-party computation for scalable federated learning," *Journal of Privacy and Security in AI*, 2022. [Online]. Available: https://arxiv.org/abs/2205.09145

[21] J. Liu, S. Wu, and F. Tramer, "Privacy-utility trade-offs in differentially private federated learning," *Neural Information Processing Systems (NeurIPS)*, 2021. [Online]. Available: https://proceedings.neurips.cc/paper/2021/file/21ab6b928e93eb72c0d85a02c68e1571-Paper.pdf

[22] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical Secure Aggregation for Federated Learning on User-Held Data," Nov. 2016, arXiv:1611.04482 [cs, stat]. [Online]. Available: http://arxiv.org/abs/1611.04482

[23] H. Hu, Z. Salcic, L. Sun, G. Dobbie, P. S. Yu, and X. Zhang, "Membership Inference Attacks on Machine Learning: A Survey," *ACM Comput. Surv.*, vol. 54, no. 11s, pp. 235:1–235:37, Sep. 2022. [Online]. Available: https://doi.org/10.1145/3523273

[24] J. Liu, J. Zhang, M. A. Jan, R. Sun, L. Liu, S. Verma, and P. Chatterjee, "A Comprehensive Privacy-Preserving Federated Learning Scheme With Secure Authentication and Aggregation for Internet of Medical Things," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 6, pp. 3282–3292, Jun. 2024, conference Name: IEEE Journal of Biomedical and Health Informatics. [Online]. Available: https://ieeexplore.ieee.org/document/10227521

[25] L. Zhao, Q. Wang, Q. Zou, Y. Zhang, and Y. Chen, "Privacy-Preserving Collaborative Deep Learning with Unreliable Participants," Oct. 2019,

arXiv:1812.10113 [cs]. [Online]. Available: http://arxiv.org/abs/1812.10113

[26] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," in *arXiv preprint arXiv:1712.07557*, 2017.

[27] X. Wang, S. Li, P. Zhang, and Y. Chen, "Knowledge distillation for lightweight and explainable intrusion detection in resource-constrained consumer devices," *IEEE Transactions on Consumer Electronics*, vol. 71, no. 1, pp. 45–55, 2025.

[28] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, and Y. Zhou, "A Hybrid Approach to Privacy-Preserving Federated Learning," in *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*, ser. AISec'19. New York, NY, USA: Association for Computing Machinery, Nov. 2019, pp. 1–11. [Online]. Available: https://doi.org/10.1145/3338501.3357370

[29] G. Caporossi and A. Taik, "Enhancing privacy in the early detection of sexual predators through federated learning and differential privacy," *arXiv preprint arXiv:2501.12537*, 2025.

[30] L. Zhang, Z. Wu, H. Xu, D. Niyato, and C. S. Hong, "Digital twin-driven federated learning for converged computing and networking at the edge," *IEEE Network*, 2024.

[31] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "A survey on federated learning systems: Vision, hype and reality for data privacy and protection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3484–3505, 2021.

[32] J. Xu, B. S. Glicksberg, C.-H. Su, P. Walker, J. Bian, and F. Wang, "Federated learning for healthcare informatics," *Journal of Healthcare Informatics Research*, vol. 5, no. 1, pp. 1–19, 2021.

[33] G. Kaissis, M. Makowski, D. Rückert, and R. Braren, "End-to-end privacy preserving deep learning on multi-institutional medical data," *Nature Machine Intelligence*, vol. 3, no. 6, pp. 473–484, 2021.

[34] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, P. Wiebe, P. Peterson, P. Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, *Array programming with NumPy*. O'Reilly Media, 2020.

[35] Faker Developers, "Faker: Python package for generating fake data," https://github.com/joke2k/faker, 2018.

[36] S. Zhang, J. Huang, and P. Li, "Analyze and improve differentially private federated learning: A model robustness perspective," *IEEE Transactions on Information Forensics and Security*, 2024.

[37] G. Amreen and A. Kanavalli, "Privacy pinnacle: Improvising healthcare data security through federated learning and blockchain framework," in *15th International Conference on Security in Computing and Communications*, 2024.

[38] Z. Cheng, Y. Liu, C. Wu, Y. Pan, L. Zhao, and X. Deng, "Decentralized iot data sharing: A blockchain-based federated learning approach with joint optimizations for efficiency and privacy," *Future Generation Computer Systems*, 2024.

[39] M. Hidayat, Y. Nakamura, and Y. Arakawa, "Enhancing efficiency in privacy-preserving federated learning for healthcare: Adaptive gaussian clipping with dft aggregator," *IEEE Access*, 2024.

[40] Q. Chen, G. Zhu, and H. Jiang, "Communication-and-energy efficient over-the-air federated learning," *IEEE Transactions on Network Science*, 2024.

[41] Z. Liu, C. Yang, Y. Ding, and H. Liang, "A lightweight and accuracy-lossless privacy-preserving method in federated learning," *IEEE Internet of Things Journal*, 2024.

[42] P. Dong, F. Zhou, and Q. Wu, "Federated transfer learning based cooperative wideband spectrum sensing with model pruning," in *IEEE/CIC International Conference on Communications in China (ICCC)*, 2024.

[43] X. Qin, H. Zhong, and X. Zhang, "Privacy preserving quantum search mechanism using grover's algorithm," *Proceedings of IEEE Quantum Computing and Networking Conference*, 2024.

[44] S. Yeom, I. Giacomelli, M. Fredrikson, and S. Jha, "Privacy risk in machine learning: Analyzing the connection to overfitting," in *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*. Oxford, United Kingdom: IEEE, 2018, pp. 268–282.

[45] R. Shokri, M. Stronati, C. Song, and V. Shmatikov, "Membership inference attacks against machine learning models," in *IEEE Symposium on Security and Privacy (SP)*, 2017.

[46] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography Conference*. Springer, 2006, pp. 265–284.

[47] B. Jayaraman and D. Evans, "Evaluating differentially private machine learning in practice," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA, USA: USENIX Association, 2019, pp. 1895–1912. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/jayaraman

[48] S. Li, E. C.-H. Ngai, and T. Voigt, "An experimental study of byzantine-robust aggregation schemes in federated learning," *IEEE Transactions on Big Data*, 2023.

[49] Q. Wu, K. Guo, and et al., "A comprehensive survey on privacy-preserving federated learning," *arXiv preprint arXiv:2003.01395*, 2020. [Online]. Available: https://arxiv.org/abs/2003.01395

[50] A. Koloskova, R. McKenna, and Z. Charles, "Gradient descent with linearly correlated noise: Theory and applications to differential privacy," *Advances in Neural Information Processing Systems*, 2023.

[51] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2007, pp. 94–103.

[52] Y. Zhang *et al.*, "Improved privacy–utility trade-offs in federated learning," *IEEE Transactions on Neural Networks and Learning Systems*, 2024, early Access.

NEO R. LEOPE (ISC CC) received the Bachelors in Information Technology from Nelson Mandela University and the B.Sc. Honours in Computer Science from the University of Pretoria, South Africa. He is currently pursuing the M.Sc. degree in Computer Science at the University of Pretoria. His research focuses on privacy-preserving federated learning, including noise-based differential privacy mechanisms, secure aggregation, and the privacy–utility trade-off in healthcare and finance.

He is employed as a Software Development Engineer at CSG International, where he works on large-scale telecom transformation projects, including customer management and billing systems for Standard Bank Mobile and Malitel. His broader research interests include cybersecurity, secure multi-party computation, privacy-aware AI, and resilient distributed systems.

JAN H. P. ELOFF was the Research Director of the SAP Research in Africa, from 2008 to 2015. From 2016 to 2021, he was the Deputy Dean Research, and the Acting Dean of the Faculty of Engineering, Built Environment and IT, University of Pretoria, South Africa, in 2022. He is currently a Full Professor in computer science with the University of Pretoria. He holds a B2 rating from the National Research Foundation in South Africa indicating that he receives considerable international recognition for his research in safeguarding platforms against societal and organizational cyber-threats. He is also a leading international scholar in conducting research on the convergence of cyber-security and AI. He has published widely in leading international journals. In 2018, he published a scholarly book on software failure investigations. He is the co-inventor of a number of patents registered in the USA. He is a member of the governing and advisory board of the International Knowledge Centre for Engineering Sciences and Technology (UNESCO(IKCEST)), China. During his research career, he represented South Africa as an Expert on Technical Committee 11 (Information Security) IFIP and was a recipient of the IFIP Silver Core and Outstanding Services Award. He also served as the South African Representative for the International Standards Organization (ISO) and as a former President of South African Institute of Computer Scientists and Information Technologists (SAICSIT). In 2017, he received a SAICSIT award recognizing him as an individual who has played a pioneering role in promoting computer science and information technology as academic disciplines in South Africa. In 2020, he received the Chancellor's Medal for Research from the University of Pretoria. He is listed as a finalist for the NSTF Lifetime Award for exemplary life-long research in cybersecurity, in 2021. He is an Associate Editor of Computers and Security, the world's leading journal for the advancement of computer security.

MOSES T. DLAMINI (PhD, CISSP, CCSP and CGRC) is a senior cybersecurity researcher in the Information and Cyber Security Centre. He has lectured on Computer Science courses with a specific focus on Information Security at the University of Pretoria, University of KwaZulu Natal, Wits University and University of Swaziland. His research and consultancy work focuses on intersection of Cybersecurity and Artificial Intelligence, Gen AI and Agentic AI; Cryptography; Cybersecurity GRC; Cybersecurity for Industrial Control Systems, OT and IoT; Cloud Security; Cybersecurity Maturity Assessment; Privacy-preserving Federated Learning, Digital Forensics, and Digital Identity of Non-Human-Identities. Dr Dlamini is passionate about technology that serves the needs of community and industry. He publishes his research work both at national and international forum.

APPENDIX A EXTENDED USE-CASE NARRATIVES AND DIAGRAMS

A. USE CASE DIAGRAMS

Figures 1 and 2 depict the use cases of FL applied to healthcare providers, aggregated at the national level. Due to space concerns, aggregation at an intermediary level, such as provincial or state, was excluded. Primary actors included patients, doctors, and healthcare facilities. These interact with secondary actors, namely healthcare systems and servers, to allow the segmentation, aggregation, and training of the model.

Within a healthcare facility, data are segmented and distributed across different data stores to preserve paired data privacy, ensuring that access to one part does not reveal sensitive patient information. Data contributed by healthcare facilities in a province were used to train the model, where noise was injected at the aggregation point to prevent inference attacks. These data can be further aggregated at the national level, as shown in the national use case diagram.

Each diagram emphasizes the importance of maintaining data privacy throughout the model training process and implicitly reflects the need to preserve data utility, ensuring that privacy mechanisms do not significantly degrade the model's performance or applicability in real-world healthcare scenarios. The next section explores the FL pipeline to better understand the various participants, how noise can be injected, and the associated objectives.

APPENDIX B EXTENDED PER-NOISE CALIBRATION NOTES

The results show that the Gaussian noise has an optimal scale of $(5, 3000)$. Its noise scale pair reflects a PUL of 0.3849 and good model accuracy of 0.3515 (MSE). Table 7 lists the lower and higher scales that resulted in either insufficient privacy protection or accuracy degradation.

Laplace noise with a scale of $(5, 3000)$ shows superior performance with the lowest PUL result of 0.3709, which is effectively balanced with the accuracy shown from the resulting MSE of 0.3209.

For the Exponential noise type, optimal performance was achieved with the scale $(5, 3000)$. This noise scale achieved a PUL of 0.3750 and minimal deviation in the MSE of 0.3210. Higher or lower scales result in an increased loss, which is an indicator of either low or excessive noise.

Similar to other noise types of the experimental noise scales, Uniform noise follows with a scale of $(5, 3000)$, which results in the best privacy–utility balance with a PUL of 0.3765 and an MSE of 0.3215. Other noise scales demonstrated increased PULs, highlighting the ineffectiveness of the scale. For example, a lower noise scale $(1, 500)$ exhibited a PUL of 0.3930, and the higher noise scale $(10, 5000)$ exhibited a PUL of 0.3940, which is a significant increase compared to the PUL of 0.3765 exhibited from the $(5, 3000)$ noise scale.

Poisson noise provides the most balanced scale $(5, 3000)$. This was reflected in the presented PUL of 0.3780, and the

accuracy was represented by an MSE of 0.3225. Noise scales higher than or lower than this result in a less favorable balance between privacy and utility. This highlights that the scale of $(5, 3000)$ is the most appropriate for providing a balanced privacy and utility trade-off.

APPENDIX C EXTENDED EXPERIMENTAL SETUP DETAILS

Noise was introduced at the client side prior to transmitting updates to the server. The study considered five statistical perturbation mechanisms, Gaussian, Laplace, Exponential, Uniform, and Poisson, alongside a no-noise baseline. Each mechanism was parameterized to reflect realistic healthcare variability. For the Gaussian mechanism, noise was drawn from $\mathcal{N}(0, \sigma^2)$, while Laplace noise was sampled from Laplace$(0, b)$. Exponential perturbations were shifted by subtracting the scale parameter to ensure zero-centered offsets. Uniform noise was sampled from $\mathcal{U}[-a, a]$, and Poisson perturbations were centered by subtracting λ. The magnitude of perturbation differed across features: the age attribute was perturbed with a scale of ± 5 years, whereas the billing attribute was perturbed with a scale of ± 3000 ZAR. This ensured that perturbations remained plausible in the healthcare context while still providing measurable privacy guarantees.

Both the clients and the central server employed identical models, defined as a two-layer neural network implemented in Keras: a fully connected hidden layer with ten ReLU activations, followed by a single sigmoid output neuron. Model training at the client side minimized a privacy–utility loss function,

$$\mathcal{L}_{\text{PU-train}} = \alpha \cdot \text{MSE}(y, \hat{y}) + (1 - \alpha) \cdot \frac{1}{\varepsilon},$$

with $\alpha = 0.1$ and $\varepsilon = 0.5$. Each selected client trained locally for one epoch before transmitting weights for aggregation. To enhance robustness against adversarial updates, global model aggregation was performed using the Krum algorithm, which selects updates closest to the majority to mitigate the influence of outliers.

Evaluation was conducted using multiple complementary metrics. Utility was quantified via MSE and MAE on both the original and noisy validation sets. Privacy was assessed using the Frobenius distance between successive global weight matrices, defined as

Frobenius Distance Calculation:

$$D_F = \sqrt{\sum (X - \hat{X})^2} \tag{16}$$

where X is the original data matrix and $\hat{X}$ is the noisy data. This metric quantified the overall magnitude of transformation induced by noise, making it particularly useful in FL to track model deviations across rounds and evaluate robustness, privacy, and utility trade-offs [50]. In addition, a post-hoc privacy–utility monitor was logged as

$$\mathcal{L}_{\text{PU-monitor}} = \text{MSE} + \varepsilon \cdot \alpha,$$

providing an interpretable measure of the balance between privacy and predictive accuracy.

Beyond quantitative analysis, the experiments also generated qualitative outputs. For each epoch and noise type, scatter plots of the validation dataset were produced, showing the transformation of original data points into their noisy counterparts. Five records were highlighted per experiment, with connector lines indicating the magnitude and direction of perturbations. These visualizations, together with CSV outputs of the full dataset and the highlighted records, provided an additional layer of interpretability regarding the privacy–utility trade-off. In this way, the experimental setup ensured a rigorous, repeatable evaluation of noise mechanisms in federated learning under realistic healthcare-inspired conditions.

A. SYNTHETIC DATASET AND EXPLORATORY ANALYSIS (FIGURES)

Table 8 presents a 5-row sample of the synthetic healthcare data used in our experiments. Values fall within realistic ranges for patient age and billing amounts (hundreds to tens of thousands of ZAR).

To assess realism prior to training, we conducted an exploratory data analysis (EDA). The Age distribution spans the full range without pathological spikes (Fig. 10). Billing shows wide dispersion and mild right-tail behavior in linear scale (Fig. 11) that becomes clearer under a log transform (Fig. 12). A boxplot summarizes variability and potential outliers (Fig. 13), while the Age–Billing scatter indicates no strong linear association (Fig. 14). Comprehensive descriptive statistics (count, mean, standard deviation, quartiles, min/max) are reported in the Appendix (Table 9). These checks indicate the dataset is statistically plausible and suitable as a testbed for the privacy–utility evaluations that follow.

B. PRIVACY-UTILITY LOSS FUNCTION (FORMULA)

Traditional federated learning optimizes purely for model accuracy using utility-based objectives such as mean squared error (MSE). However, these formulations ignore the explicit cost of privacy preservation. To capture the inherent trade-off, we introduce a Privacy–Utility Loss (PUL) function that incorporates both prediction error and privacy guarantees.

Formally, let the prediction error be measured by the mean squared error:

$$\mathcal{L}_{\text{utility}} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2,$$

and let the privacy cost be modeled as the inverse of the differential privacy parameter ε:

$$\mathcal{L}_{\text{privacy}} = \frac{1}{\varepsilon}.$$

The combined Privacy–Utility Loss is then expressed as

$$\mathcal{L}_{PU} = \alpha \cdot \mathcal{L}_{\text{utility}} + (1 - \alpha) \cdot \mathcal{L}_{\text{privacy}},$$

where $\alpha \in [0, 1]$ is a weighting parameter that tunes the balance between accuracy and privacy. Higher values of α

prioritize predictive accuracy, while lower values increase emphasis on privacy.

This formulation enables explicit exploration of the privacy–utility trade-off, which is often implicit in standard FL training. Similar approaches to balancing error and privacy have been discussed in prior work on differentially private machine learning [8], [36], [41], [47].

C. DIFFERENTIAL PRIVACY FORMALISM (BACKGROUND)

Differential privacy (DP) provides a formal guarantee of privacy by bounding the effect of any single individual record on the output of a randomized algorithm. Formally, a mechanism $\mathcal{M}$ is said to satisfy (ε, δ)-differential privacy if for any two neighboring datasets D_1 and D_2 differing in one record, and for all measurable subsets S of the output space:

$$\Pr[\mathcal{M}(D_1) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D_2) \in S] + \delta.$$

Here, ε is the privacy budget (smaller values indicate stronger privacy) and δ is the probability of failure of the privacy guarantee [9].

The noise mechanisms adopted in this study can be interpreted under this framework. The Laplace mechanism achieves pure ε-DP with $\delta = 0$, where the scale parameter b calibrates the noise to the sensitivity of the query. The Gaussian mechanism satisfies (ε, δ)-DP, with the standard deviation σ linked to both ε and δ [8]. The Exponential mechanism provides ε-DP guarantees by probabilistically favoring outputs of higher utility [51]. Poisson and uniform perturbations, while not standard DP mechanisms, are frequently considered in comparative studies of noise distributions for federated learning [49], [52].

Although the experiments in this paper calibrated noise based on scale parameters (e.g., ± 5 years for age, ± 3000 ZAR for billing), these settings can be interpreted as approximations of (ε, δ) values under standard DP analysis. Future work will integrate a tighter DP accountant to provide explicit numerical guarantees.

APPENDIX D ADDITIONAL FIGURES

The following pages present the key figures referenced throughout the paper, summarizing the experimental findings and supporting the discussions provided in the text.

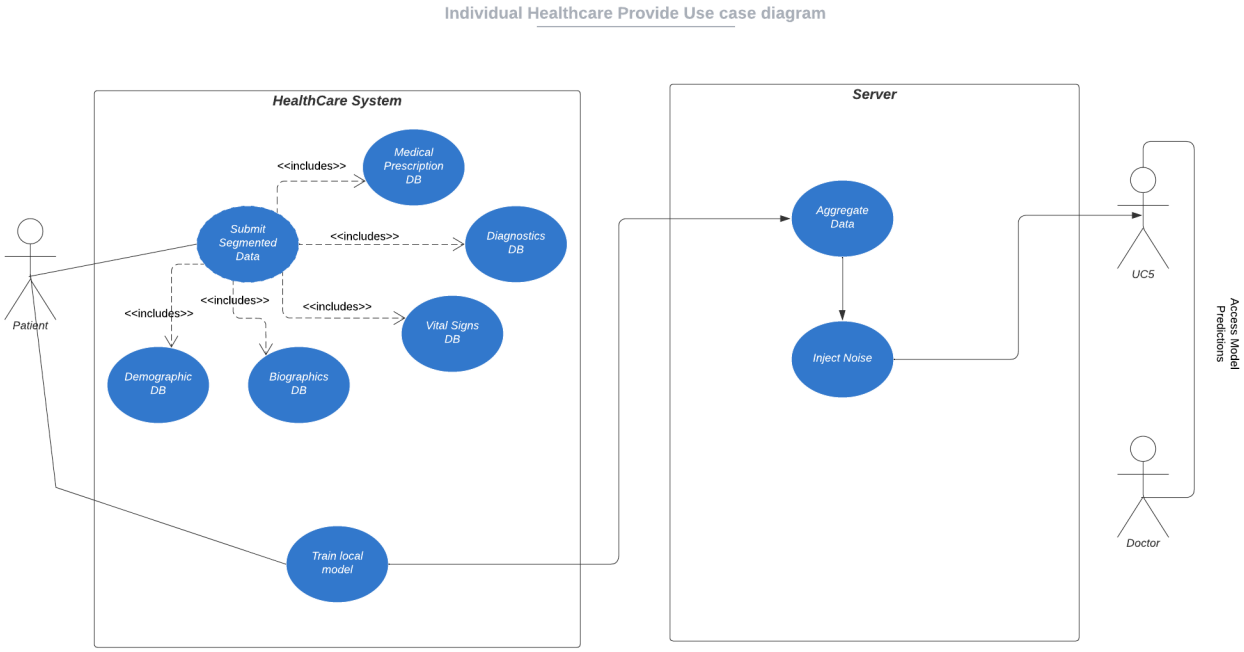

FIGURE 1: **Healthcare Facility Use Case.** This diagram illustrates the federated learning setup at a single healthcare facility.

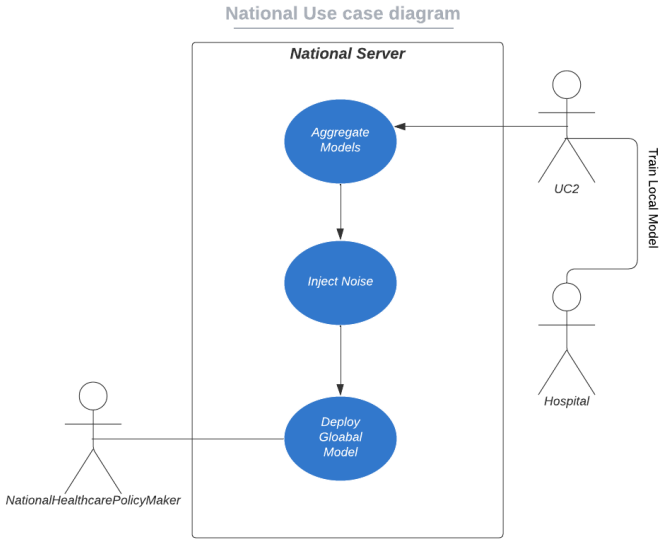

FIGURE 2: **National Use Case.** This diagram illustrates aggregation across multiple facilities at the national level.

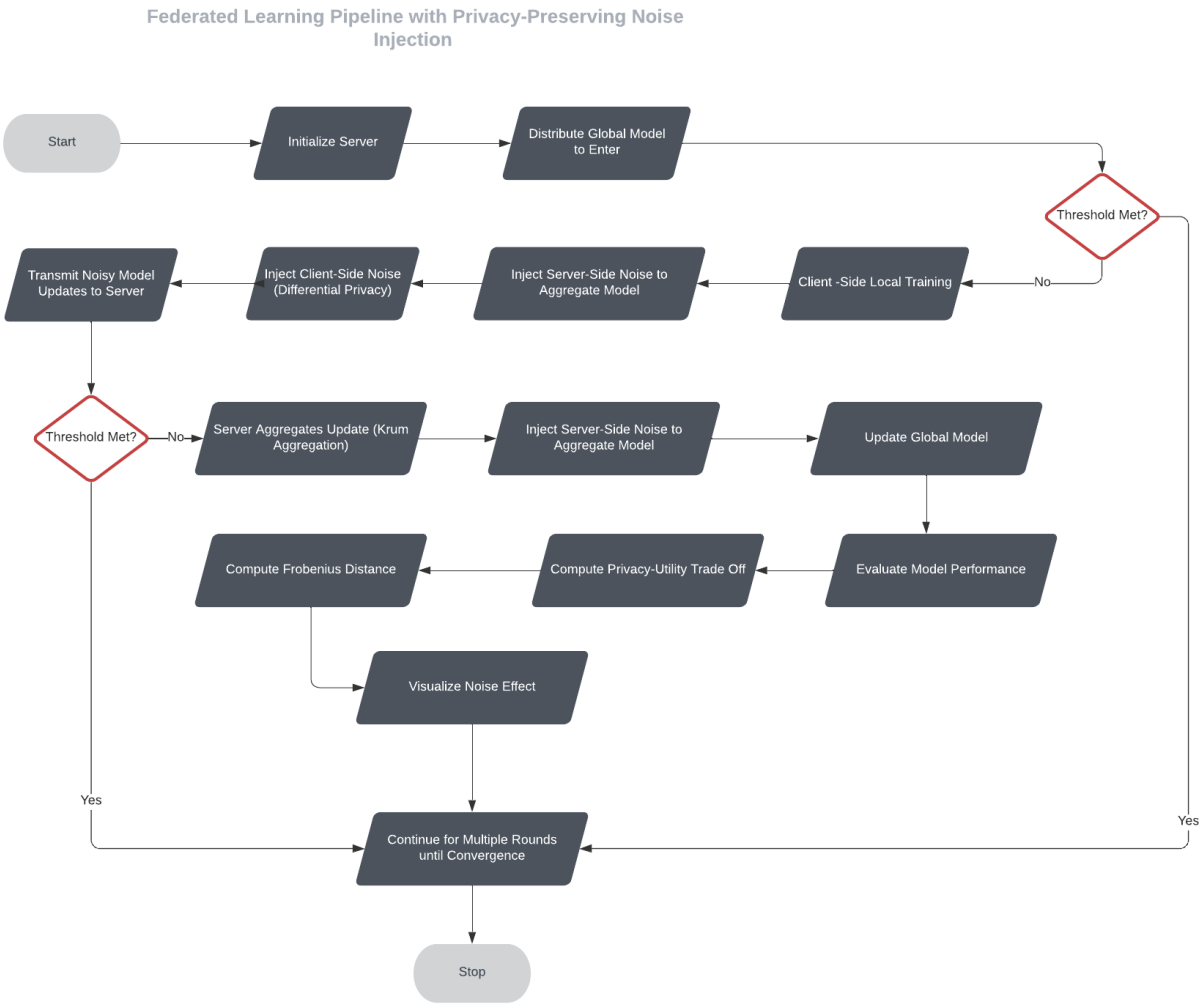

FIGURE 3: Federated learning pipeline with privacy-preserving noise injection. Blocks: (1) *Initialize Server* creates the global model $W^{(0)}$; (2) *Distribute Global Model to Clients*[1] sends $W^{(t)}$ to selected clients; (3) *Client-Side Local Training* performs one local epoch; (4A) *Inject Client-Side Noise (Differential Privacy)* adds calibrated noise at the **feature level** before or during local training; (4B) *Inject Server-Side Noise to Aggregate Model* denotes an *optional* gradient/aggregate perturbation (not used in our experiments, shown for completeness); (5) *Transmit (Noisy) Updates to Server*; (6) *Threshold Met?* checks the *aggregation threshold* k; if fewer than k clients responded, the round waits/aborts; (7) *Server Aggregates Update (Krum)* robustly aggregates client updates; (8) *Update Global Model* to $W^{(t+1)}$; (9) *Evaluate Model Performance* on held-out validation (MSE, MAE); (10) *Compute Privacy–Utility Trade-Off* (PUL monitor); (11) *Compute Frobenius Distance* $\|W^{(t+1)} - W^{(t)}\|_F$ to track perturbation; (12) *Visualize Noise Effect*; (13) *Continue for Multiple Rounds until Convergence*. In this paper we enable (4A) only; (4B) is a design alternative left to future work.

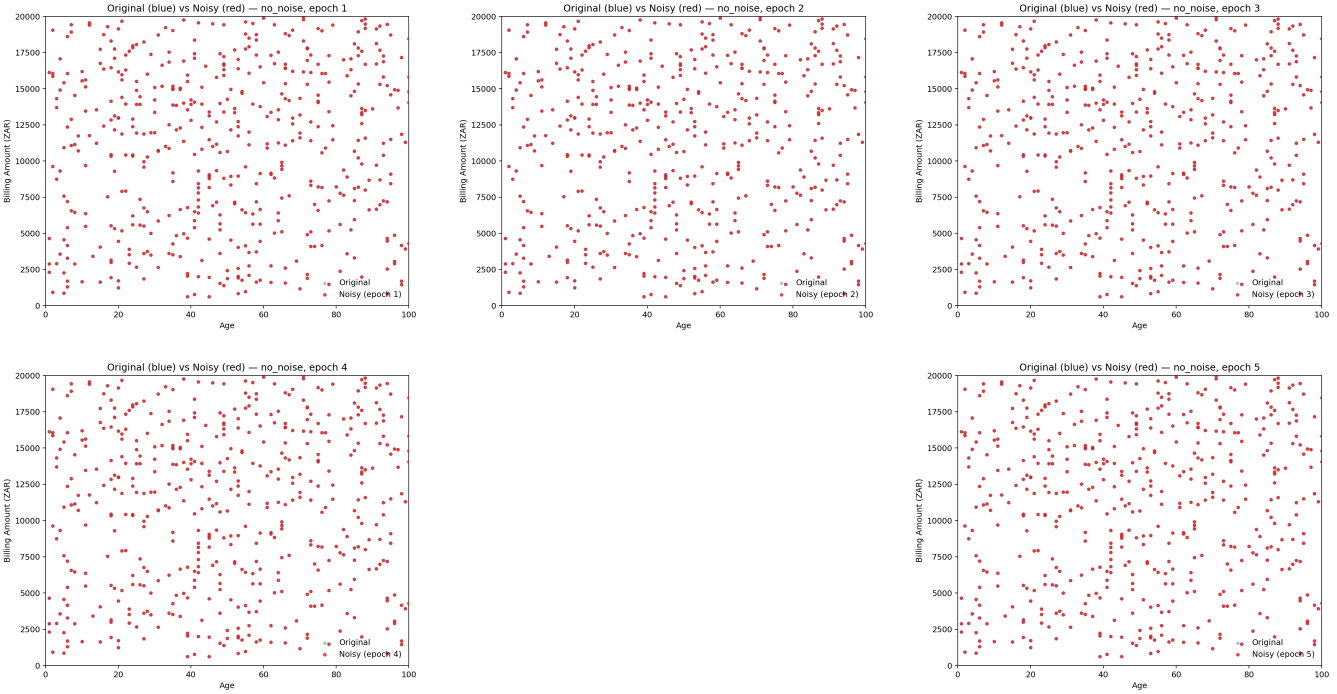

FIGURE 4: Visualization of Baseline (No Noise) Experiments, Epochs 1–5.

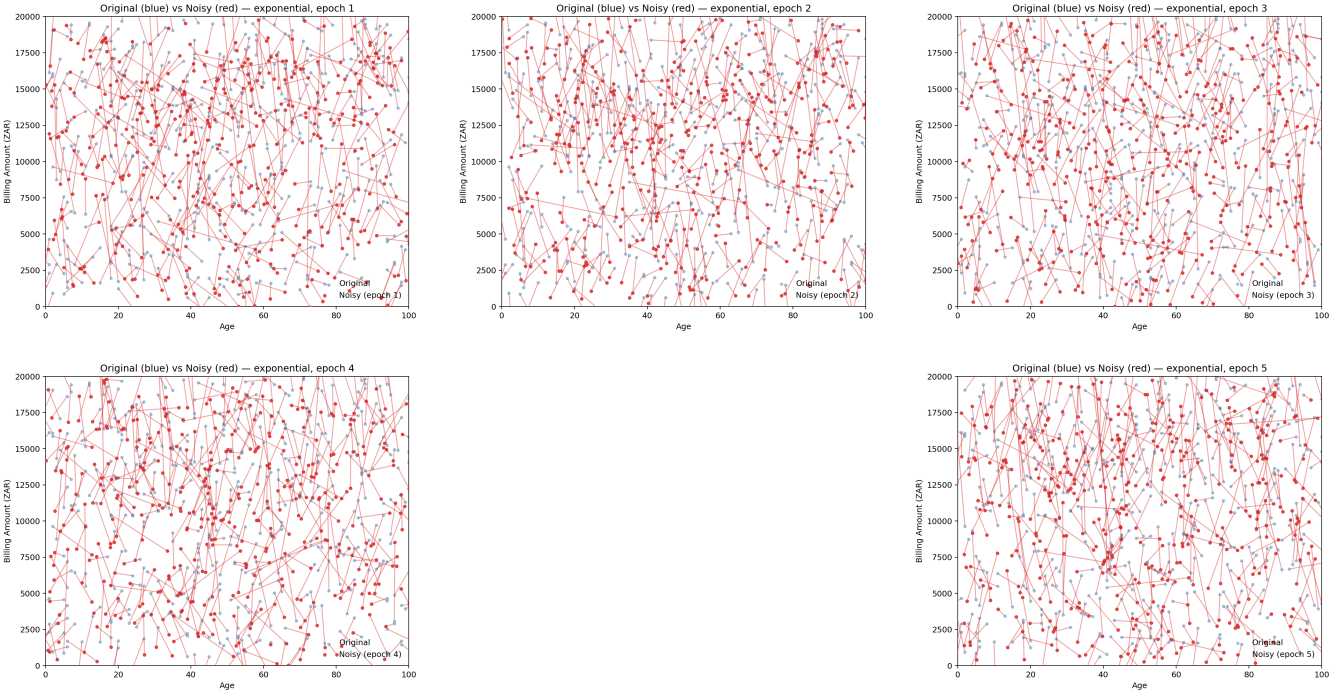

FIGURE 5: Visualization of Exponential Noise Experiments, Epochs 1–5.

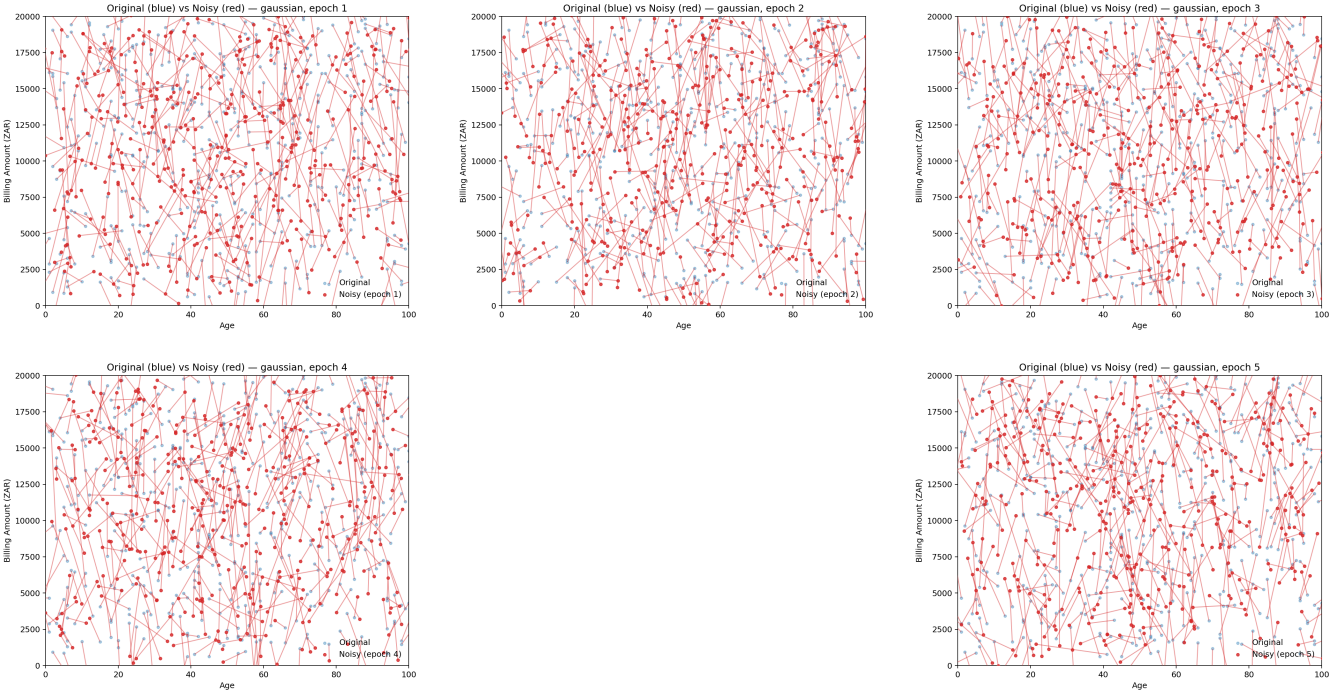

FIGURE 6: Visualization of Gaussian Noise Experiments, Epochs 1–5.

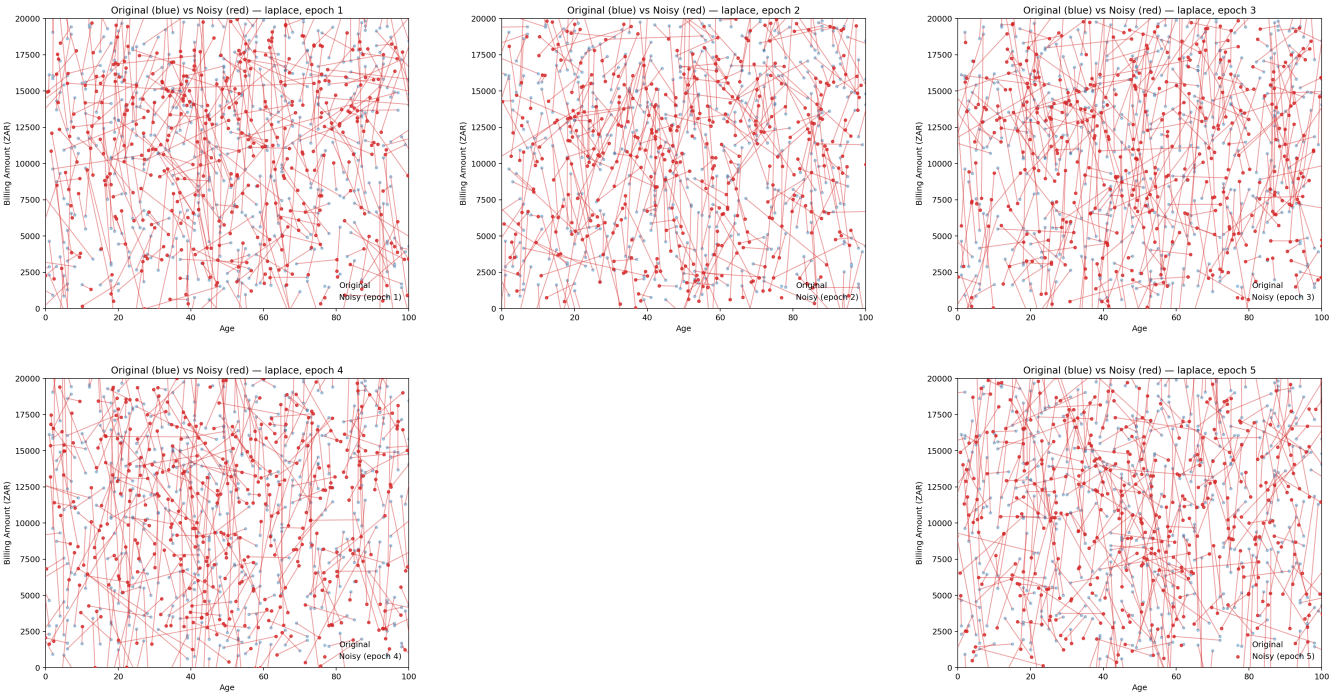

FIGURE 7: Visualization of Laplace Noise Experiments, Epochs 1–5.

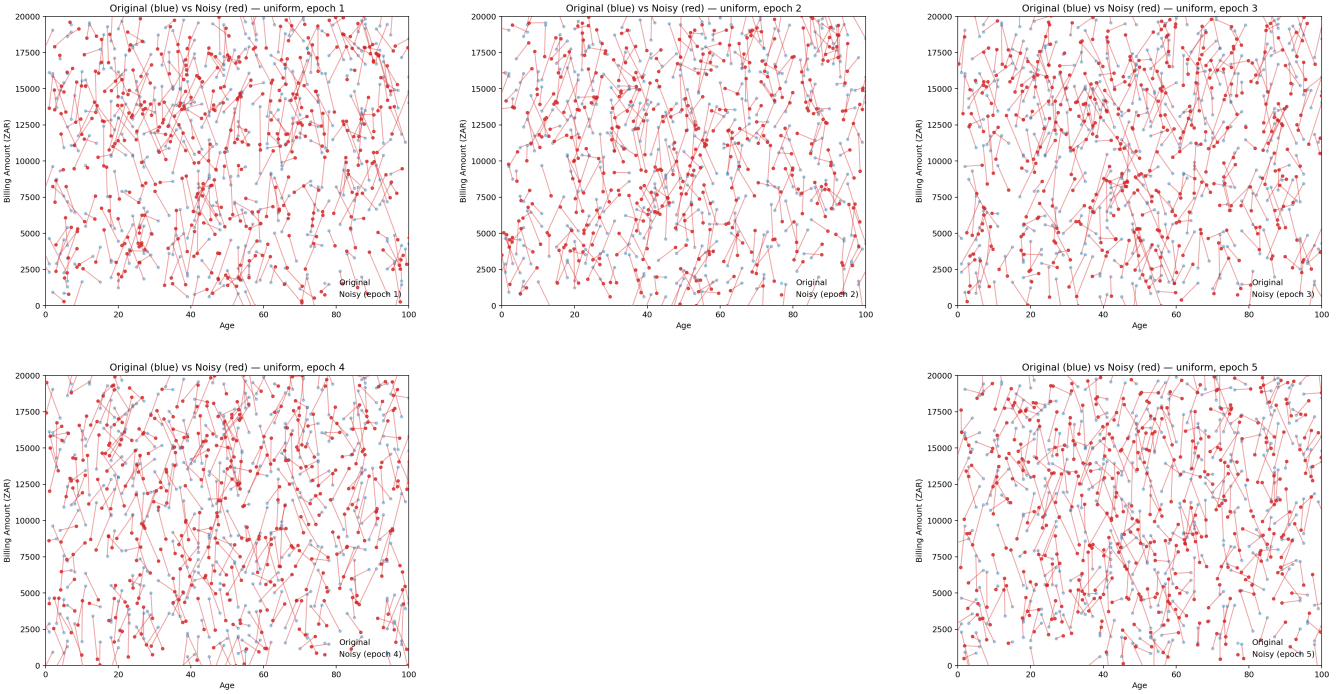

FIGURE 8: Visualization of Uniform Noise Experiments, Epochs 1–5.

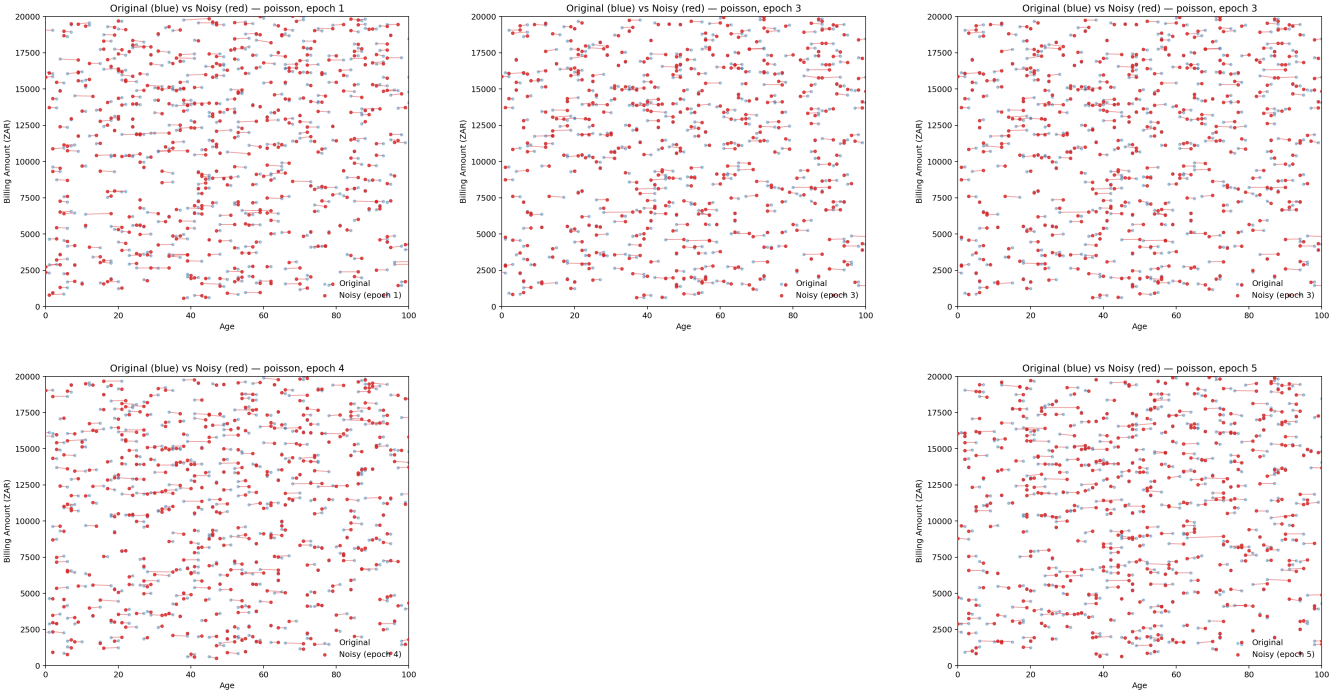

FIGURE 9: Visualization of Poisson Noise Experiments, Epochs 1–5.

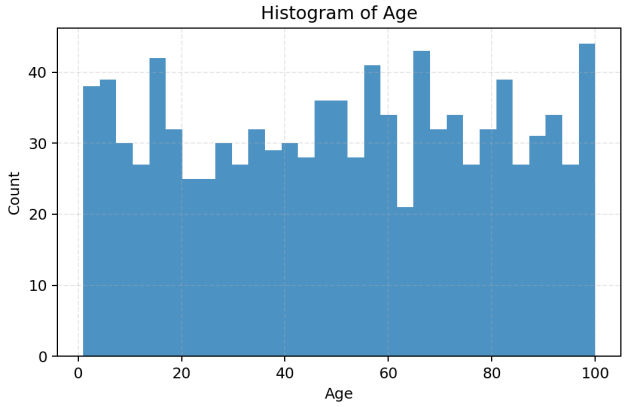

FIGURE 10: Histogram of Age.

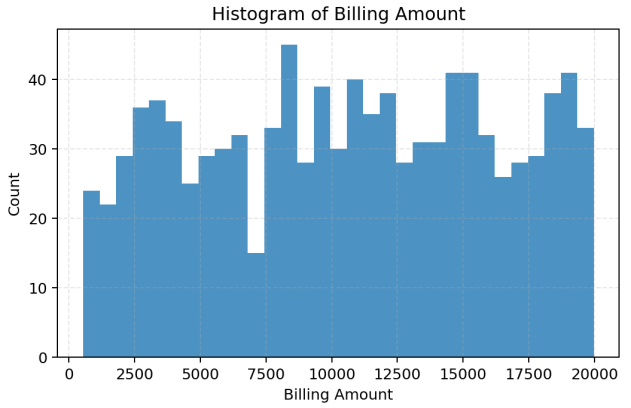

FIGURE 11: Histogram of Billing Amount (linear scale).

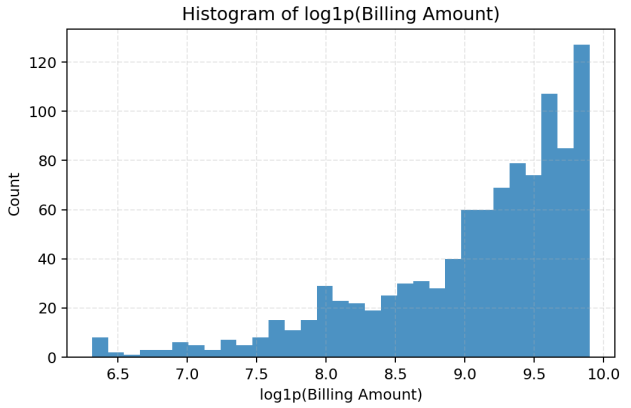

FIGURE 12: Histogram of Billing Amount (log1p scale) illustrating tail behavior.

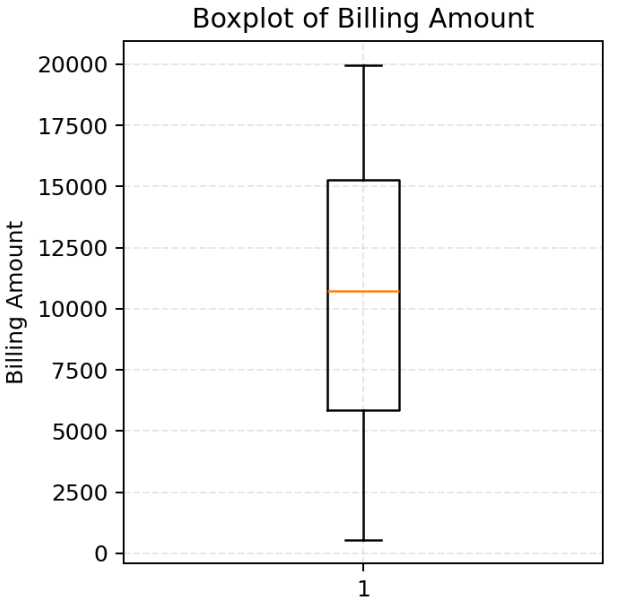

FIGURE 13: Boxplot of Billing Amount with 1.5×IQR fences.

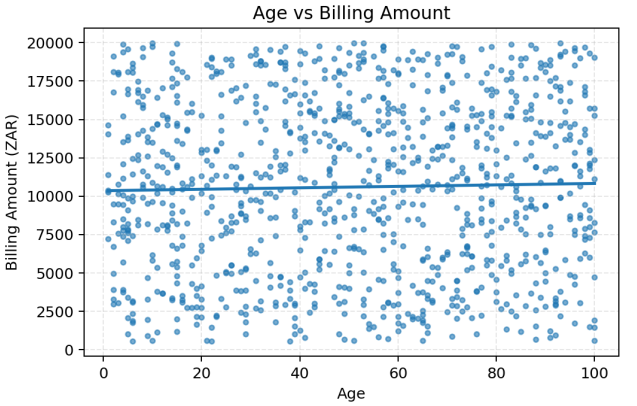

FIGURE 14: Age vs. Billing Amount.

APPENDIX E ADDITIONAL TABLES

The following pages present the key tables referenced throughout the paper, summarizing the experimental findings and supporting the discussions provided in the text.

TABLE 7: Comparison of Noise Types and Scales on Model Accuracy and Privacy

Noise Type	Scale (Age, Billing)	MSE (Original)	MSE (Noisy)	PUL
Gaussian	(1, 500)	0.3560	0.3560	0.3990
Gaussian	(5, 3000)	0.3560	0.3515	0.3849
Gaussian	(10, 5000)	0.3209	0.3209	0.3912
Laplace	(1, 500)	0.3209	0.3256	0.3919
Laplace	(5, 3000)	0.3209	0.3209	0.3709
Laplace	(10, 5000)	0.3560	0.3490	0.3849
Exponential	(1, 500)	0.3209	0.3250	0.3925
Exponential	(5, 3000)	0.3209	0.3210	0.3750
Exponential	(10, 5000)	0.3209	0.3260	0.3900
Uniform	(1, 500)	0.3209	0.3265	0.3930
Uniform	(5, 3000)	0.3209	0.3215	0.3765
Uniform	(10, 5000)	0.3209	0.3270	0.3940
Poisson	(1, 500)	0.3209	0.3300	0.3965
Poisson	(5, 3000)	0.3209	0.3225	0.3780
Poisson	(10, 5000)	0.3209	0.3290	0.3950
No Noise (Baseline)	(0, 0)	0.3209	0.3209	0.3900

TABLE 8: Sample of Synthetic Healthcare Dataset

Name	Age	Gender	Hospital	Billing Amount (ZAR)
John Smith	45	M	Pretoria General Hospital	12,540
Mary Johnson	32	F	Cape Town General Hospital	7,890
Robert Williams	67	M	Johannesburg General Hospital	18,230
Patricia Brown	54	F	Durban General Hospital	11,475
Michael Davis	28	M	Bloemfontein General Hospital	6,320

TABLE 9: Descriptive Statistics of Numeric Features

Feature	Count	Mean	Std	Min	P25	P50	P75	Max	Skew
Age	1,000	50.908	29.142	1	26.00	52.00	76.00	100	−0.035
Billing Amount	1,000	10,596.86	5,526.42	551	5,847.75	10,727.50	15,262.00	19,966	−0.057

TABLE 10: Privacy and Utility Summary (Final Epoch). Proxy rows study perturbations without formal DP guarantees; DP rows report formal accounting.

Setting	Formal?	Mechanism	ϵ @ δ=10^{-5}	Val MSE	Val MAE	TPR	FPR	Adv
Baseline (no noise)	No (proxy)	—	—	0.3385	0.5020	0.591	0.572	0.019
Gaussian feature-noise	No (proxy)	Feature perturbation	—	0.3539	0.5198	0.590	0.552	0.038
DP–SGD (Gaussian)	Yes	Clip C=1.0, σ=1.2	4.216	0.3359	0.5065	0.584	0.564	0.020
Gaussian LDP (ϵ=4)	Yes	One-shot LDP	4.0 (target)	0.3148	0.4800	0.605	0.572	0.033

TABLE 11: Gaussian LDP at ϵ=4, δ=10^{-5}: variance across random seeds (5 runs). Metrics from the final epoch of each run. Mean±SD shown in the last row.

Seed	Val MSE	Val MAE	PU Loss	Frobenius	TPR	FPR	Adv
1	0.3463	0.5089	0.3963	3.614	0.629	0.492	0.137
2	0.3363	0.5039	0.3863	2.071	0.615	0.532	0.083
3	0.3073	0.4786	0.3573	3.033	0.621	0.566	0.055
4	0.3387	0.5052	0.3887	3.802	0.620	0.540	0.080
5	0.3157	0.4857	0.3657	3.351	0.611	0.570	0.041
Mean±SD	**0.3307±0.0160**	**0.4981±0.0145**	**0.3860±0.0159**	**3.174±0.686**	**0.619±0.0068**	**0.540±0.031**	**0.079±0.0367**

TABLE 12: Gaussian LDP sensitivity to ϵ (normalized σ from the accountant; δ=10^{-5}). One run per setting, final-epoch metrics.

ϵ	Calibrated σ	Val MSE	Val MAE	PU Loss	Frobenius	TPR	FPR	Adv
2.0	3.4258	0.3498	0.5190	0.3998	3.308	0.637	0.500	0.137
3.0	2.2839	0.3528	0.5169	0.4028	0.001	0.597	0.484	0.113
4.0	1.7129	0.3271	0.4916	0.3771	2.835	0.612	0.546	0.066
6.0	1.1419	0.3351	0.4990	0.3851	1.734	0.610	0.530	0.080
8.0	0.8564	0.3216	0.4887	0.3716	2.868	0.603	0.578	0.025

TABLE 13: Non-IID stress test (age-skewed CSV), Gaussian LDP at ϵ=4, δ=10^{-5}.

Setting	Val MSE	Val MAE	PU Loss	Frobenius	TPR	FPR	Adv
Age-skewed (non-IID)	0.3387	0.5079	0.3887	1.235	0.604	0.524	0.080

TABLE 14: Baseline (No Noise) Results Across Epochs

Epoch	MSE	MAE	PU Loss	Frobenius Distance
1	0.3560	0.5176	0.4060	N/A
2	0.3560	0.5176	0.4060	0.0000
3	0.3560	0.5176	0.4060	2.9188
4	0.3560	0.5176	0.4060	2.8784
5	0.3209	0.4824	0.3709	3.8801

TABLE 15: Gaussian Noise Results Across Epochs

Epoch	MSE	MAE	PU Loss	Frobenius Distance
1	0.3525	0.5146	0.4026	N/A
2	0.3526	0.5147	0.4026	0.0000
3	0.3526	0.5147	0.4026	2.9188
4	0.3488	0.5111	0.4026	2.8785
5	0.3247	0.4868	0.3733	3.8800

TABLE 16: Laplace Noise Results Across Epochs

Epoch	MSE	MAE	PU Loss	Frobenius Distance
1	0.3526	0.5147	0.4026	N/A
2	0.3515	0.5136	0.4026	0.0000
3	0.3526	0.5147	0.4026	2.9188
4	0.3446	0.5065	0.4026	2.8784
5	0.3213	0.4834	0.3733	3.8799

TABLE 17: Exponential Noise Results Across Epochs

Epoch	MSE	MAE	PU Loss	Frobenius Distance
1	0.3524	0.5145	0.4026	N/A
2	0.3523	0.5143	0.4026	0.0000
3	0.3526	0.5146	0.4026	2.9188
4	0.3507	0.5128	0.4026	2.8784
5	0.3228	0.4854	0.3733	3.8801

TABLE 18: Poisson Noise Results Across Epochs

Epoch	MSE	MAE	PU Loss	Frobenius Distance
1	0.3456	0.5089	0.3956	N/A
2	0.3456	0.5089	0.3956	2.3904
3	0.3456	0.5089	0.3956	2.9188
4	0.3456	0.5089	0.3956	2.8784
5	0.3278	0.4911	0.3778	3.8801

TABLE 19: Uniform Noise Results Across Epochs

Epoch	MSE	MAE	PU Loss	Frobenius Distance
1	0.3526	0.5147	0.4026	N/A
2	0.3544	0.5165	0.4026	0.0000
3	0.3526	0.5147	0.4026	2.9187
4	0.3485	0.5106	0.4026	2.8788
5	0.3230	0.4850	0.3733	3.8801