Client-Level Differential Privacy via Federated Distributionally Robust Optimization

Yan Yan
yan.yan1@wsu.edu
Washington State University
Pullman, Washington, USA

Zhishuai Guo
zguo@niu.edu
Northern Illinois University
DeKalb, Illinois, USA

Mingrui Liu
mingruil@gmu.edu
George Mason University
Fairfax, Virginia, USA

Ping Liu
pingl@unr.edu
University of Nevada Reno
Reno, Nevada, USA

Abstract

Federated learning (FL) faces a fundamental tension between client-level privacy protection and robustness under heterogeneous clients. Differential privacy (DP) provides formal privacy guarantees but is often enforced through repeated noise injection, which can significantly degrade utility, while federated distributionally robust optimization (FDRO) improves robustness by prioritizing hard-to-fit clients. However, its relation to client-level DP remains unclear. In this paper, we establish a precise client-level characterization of when KL-regularized FDRO instantiations coincide with the exponential mechanism (EM), thereby yielding formal client-level DP guarantees. We further characterize how the privacy parameter governs the concentration of probability mass around worst-case clients, revealing a continuous spectrum between uniform averaging and worst-client emphasis. Our results provide a unified perspective on privacy and robustness in FL. Experiments on CIFAR-10 federated benchmarks validate our theoretical findings and demonstrate a robustness trade-off: moderate bias improves worst-client performance without harming global accuracy, while excessive bias degrades broader tail robustness.

CCS Concepts

• **Computing methodologies** → **Machine learning**.

Keywords

Differential Privacy, Federated Learning, Distributionally Robust Optimization

ACM Reference Format:
Yan Yan, Zhishuai Guo, Mingrui Liu, and Ping Liu. 2018. Client-Level Differential Privacy via Federated Distributionally Robust Optimization. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Conference acronym 'XX, Woodstock, NY
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Federated learning (FL) [13, 16] enables collaborative model training across distributed clients without requiring the sharing of raw data, making it a natural choice for privacy-sensitive applications such as mobile devices [9], healthcare [22] and cross-organizational data analysis [12]. While FL reduces direct data exposure, protecting client-level information during training remains a fundamental challenge, particularly under heterogeneous data distributions and uneven client contributions.

Therefore, to deal with the privacy in FL, differential privacy (DP) has been proposed and studied as a standard tool for providing formal privacy guarantees in federated systems [1, 17], typically through noise injection mechanisms applied to model updates or aggregation procedures. There are three main ways, i.e., (i) gradient perturbations [7, 17], (ii) output perturbations [23, 29], and (iii) objective perturbations [25]. However, existing DP-FL methods often rely on repeated per-iteration perturbations [17, 27, 28], which can significantly degrade utility and exacerbate the tension between privacy, robustness, and efficiency in large-scale federated optimization [12, 28].

This tension becomes particularly pronounced in the presence of client heterogeneity, where data distributions, sample sizes and learning difficulties vary significantly across clients. Federated distributionally robust optimization (FDRO) has recently emerged as a principled framework for addressing this challenge by explicitly prioritizing worst-case or hard-to-fit clients. This explicit emphasis on client-level differences also makes FDRO a natural lens for examining how privacy mechanisms, particularly those that introduce randomness at the client level, reshape robustness objectives and optimization behavior in federated learning.

A natural connection between DP and robustness arises through the *exponential mechanism* (EM) [5, 18]. It selects outcomes with probability biased toward higher utility, where utility typically reflects per-client losses or risks in federated optimization, while preserving privacy guarantees [6]. In centralized settings, this EM has been shown to admit a distributionally robust interpretation [19, 26], linking privacy constraints to uncertainty sets defined by divergence measures. In FL settings, however, the utility function typically decomposes over clients, often reflecting per-client losses or risks, and the EM induces a nontrivial sampling distribution over clients rather than a single worst-case choice.

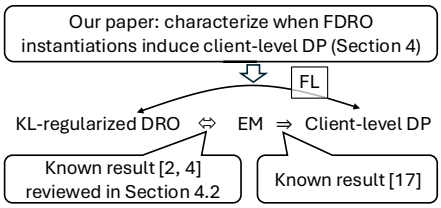

Figure 1: Roadmap of this paper: We analyze the exponential mechanism (EM) in federated learning by first characterizing its neighborhood mass concentration (Section 4.1), then interpreting the induced client reweighting as a soft KL-regularized FDRO objective (Section 4.2), and finally establishing formal client-level DP guarantees (Section 4.3).

Despite the frequent implicit use of the exponential mechanism in privacy-preserving FL algorithms, there is currently limited understanding of (i) how probability mass under the exponential mechanism concentrates around the worst-case client and its neighborhood, (ii) how this concentration depends on the privacy parameter ε, sensitivity τ, and inter-client performance gaps, and (iii) how these factors shape the effective robustness profile of the resulting model. This lack of characterization obscures the precise relationship between DP and FDRO in FL settings, and limits our ability to reason about the trade-offs between privacy strength, robustness to client heterogeneity and optimization efficiency.

In this paper, we focus on a systematic characterization of EM in FL settings through the lens of distributionally robust optimization (DRO). Our goal is to analyze how privacy-induced randomization reshapes the effective client weighting implicit in KL-regularized FDRO objectives. We (i) study the concentration behavior of EM over clients, (ii) quantify how probability mass accumulates around the worst-case client and its local neighborhood, and (iii) examine how this behavior varies as a function of the privacy parameter ε, sensitivity τ and inter-client performance gaps. This perspective reveals that privacy constraints induce a continuous spectrum between uniform client treatment and worst-case emphasis, which offers a principled interpretation of privacy–robustness trade-offs in FL. By clarifying this relationship, our analysis provides conceptual guidance that may inform the design of more efficient DP-aware FDRO methods. In particular, it suggests alternatives to repeated per-iteration perturbations, which we leave as an important direction for future work. The main contributions of this paper include:

- We provide a client-level, gap-sensitive characterization of EM in FL, showing when KL-regularized FDRO induces formal client-level DP guarantees.
- We characterize how probability mass under EM concentrates around the worst client and its loss-gap–based neighborhood, explicitly quantifying the dependence on the privacy parameter ε, sensitivity τ, and inter-client loss gaps, and revealing a continuous spectrum between uniform averaging and worst-client emphasis.
- We interpret EM as inducing a soft gap-sensitive FDRO objective that smoothly interpolates between average-risk minimization and worst-client optimization.
- We prove pure $(\varepsilon, 0)$ client-level DP guarantees for EM-based client selection and analyze how practical loss approximation leads to a controlled degradation of the privacy guarantee.
- Experiments on CIFAR-10 federated benchmarks validate the predicted privacy-robustness trade-off, showing that moderate bias improves worst-client performance without harming global accuracy, while excessive bias degrades tail robustness.

2 Related Work

FL with Heterogeneous Clients. FL enables collaborative model training across decentralized edge devices without exchanging raw data [16]. While the foundational FedAvg algorithm performs well in IID settings, its performance degrades significantly under statistical heterogeneity (non-IID data) [12, 15]. To address this issue, some approaches adopt distributionally robust optimization (DRO) at the client level to adaptively adjust aggregation weights and improve worst-case performance and fairness [3, 8, 11]. However, these methods primarily focus on optimization utility and typically overlook formal privacy guarantees.

Differential Privacy (DP). DP has emerged as the dominant formal framework for limiting information leakage in FL [1, 17]. Most DP-FL methods enforce privacy by adding carefully calibrated noise to client updates or to the aggregated global model, thereby bounding the influence of any single data record or client [17, 27, 28]. However, such noise-based mechanisms often lead to significant performance degradation in heterogeneous (non-IID) environments.

Exponential Mechanism (EM). The Exponential Mechanism (EM) [18] is a fundamental DP algorithm for selecting a high-utility outcome while preserving privacy. Instead of adding additive noise (as in Laplace or Gaussian mechanisms), EM samples outcomes from a probability distribution proportional to a utility function. Importantly, EM has been shown to be equivalent to solving a risk minimization problem with KL-regularization [19]. This interpretation establishes a natural link between privacy and robustness: entropy regularization prevents the mechanism from over-concentrating on any single data point (ensuring privacy), while still favoring high-utility outcomes (promoting robustness).

Distributionally Robust Optimization (DRO). DRO seeks models that perform well under the worst-case distribution within a prescribed uncertainty set [20, 21, 24]. In the context of FL, federated and DRO methods address client heterogeneity by optimizing for the worst-case client or group to improve generalization and fairness [3, 8, 11, 20]. These frameworks frequently employ mechanisms such as KL-regularization or Conditional Value at Risk (CVaR) constraints. In particular, DRO with KL-regularization is equivalent to a log-sum-exp formulation [8]. However, standard DRO frameworks do not inherently provide formal privacy guarantees and may inadvertently overemphasize or memorize outlying clients.

Bridging the Gap: DP and DRO in FL. There exists a fundamental tension between DRO and DP. Because DRO tends to emphasize hard or outlying samples to improve worst-case performance, while DP limits the influence of any single data point or client to ensure stability and privacy [17]. Recent theoretical studies have revealed

a deep connection between these two paradigms: in particular, KL-regularized DRO has been shown to be mathematically equivalent to the EM in DP [19].

Despite this elegant duality, its algorithmic implications have so far been only lightly explored in the context of FL, where existing methods typically treat robustness to non-IID data and privacy protection as separate objectives. This suggests an opportunity to develop FL algorithms that explicitly leverage the DP–DRO connection to jointly address statistical heterogeneity and privacy, potentially reducing the reliance on aggressive noise injection that often leads to substantial utility loss.

3 Notations and Problem Setup

3.1 Notations for Federated Client Model

We consider a standard FL setting with a fixed population of n clients, indexed by $\mathcal{I} = \{1, 2, ..., n\}$. Each client $i \in \mathcal{I}$ is associated with a local data distribution, denoted by $\mathcal{D}_i$ over the input-output space $\mathcal{X} \times \mathcal{Y}$. Let $\ell(\theta; x, y)$ denote a per-sample loss function parameterized by a model $\theta \in \Theta \subseteq \mathbb{R}^d$. Denote by $\text{Simplex}_n := \{q \in \mathbb{R}_+^n : \sum_{i=1}^n q_i = 1\}$ the probability simplex over n elements.

Client-Level Loss. For a fixed model parameter θ, the population risk of client i is defined as

$$L_i(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}_i}[\ell(\theta; x, y)].$$

Throughout this work, we adopt a client-level abstraction: for a given θ, each client is represented solely by its scalar loss value $L_i(\theta)$. When the dependence on θ is clear from context, we write L_i for brevity. This abstraction reduces the federated system to a finite set of scalar losses $\{L_i\}_{i=1}^n$, and meanwhile isolates the effect of client heterogeneity and allows us to analyze how privacy mechanisms act directly on client-level quantities, independently of the underlying optimization dynamics.

Worst Client and Client Loss Gaps. To quantify client heterogeneity in a way compatible with randomized mechanism, we define the worst-case client w.r.t. loss as $i^* \in \arg\max_{i\in\mathcal{I}} L_i$. Then, to quantify the client heterogeneity relative to the worst case, we define the client loss gap for each client i as $\Delta_i := L_{i^*} - L_i \geq 0$. In particular, the minimum non-zero gap is defined as $\Delta_{\min} := \min_{i\neq i^*} \Delta_i$, which captures how sharply separated the worst client from the rest of the population. These gap quantities will play a central role in characterizing how probability mass concentrates under privacy-induced randomization mechanisms.

3.2 Exponential Mechanism Over Clients

We now introduce the exponential mechanism (EM) as a client-level randomization mechanism acting directly on the set of federated clients.

Client-level Scoring Function. Given a fixed model parameter θ, recall that each client $i \in \mathcal{I}$ is represented by its loss value $L_i(\theta)$. We define a client-level score function $s_i := L_i$, so that higher scores correspond to larger losses. We assume the score function has bounded sensitivity τ, i.e., $\max_{S\sim S'} |s_i(S) - s_i(S')| \leq \tau$, where S and S' are two local datasets of client i of size m that differ in at most one sample, and S consists of samples drawn from the client distribution $\mathcal{D}_i$. This choice of score function is standard in EM-based constructions [18], and allows the mechanism to favor high-loss (i.e., worst-case) clients through appropriate scaling.

Exponential Mechanism over Clients. For a privacy parameter $\varepsilon > 0$, the EM induces a probability distribution over clients given by

$$p_i(\varepsilon) := \frac{\exp(\varepsilon \cdot s_i/(2\tau))}{\sum_{j=1}^n \exp(\varepsilon \cdot s_j/(2\tau)))} = \frac{\exp(\varepsilon \cdot L_i/(2\tau))}{\sum_{j=1}^n \exp(\varepsilon \cdot L_j/(2\tau))},\ i \in \mathcal{I}, \tag{1}$$

where τ denotes the sensitivity parameter. We denote by $\boldsymbol{p}(\varepsilon) = (p_1(\varepsilon), ..., p_n(\varepsilon))$ the resulting client distribution. The EM defines a soft reweighting over clients, where the relative importance of client i depends exponentially on its loss value L_i. By defining the log-partition function

$$Z(\varepsilon) := \sum_{j=1}^n \exp(\varepsilon L_j/(2\tau)),\ \psi(\varepsilon) := \log(Z(\epsilon)),$$

the induced distribution can be written compactly as

$$p_i(\varepsilon) = \exp(\varepsilon L_i/(2\tau) - \psi(\varepsilon)).$$

Relative Weighting and Pairwise Ratios. A key property of the EM is that relative probabilities depend only on loss gaps. For any two clients $i, j \in \mathcal{I}$,

$$\frac{p_i(\varepsilon)}{p_j(\varepsilon)} = \exp(\varepsilon(L_i - L_j)/(2\tau)) = \exp(\varepsilon(\Delta_j - \Delta_i)/(2\tau)).$$

In particular, relative to the worst client i^*, we have

$$\frac{p_i(\varepsilon)}{p_{i^*}(\varepsilon)} = \exp(-\varepsilon\Delta_i/(2\tau)),\ i \in \mathcal{I}, \tag{2}$$

which shows that the entire client distribution is determined by the set of loss gaps $\{\Delta_i\}_{i=1}^n$ and the privacy parameter ε.

Client-Level DP.. We adopt the standard notion of client-level DP in FL. Let $S = (S_1, \ldots, S_n)$ and $S' = (S'_1, \ldots, S'_n)$ be two neighboring federated datasets that differ in exactly one client dataset, i.e., there exists an index i such that $S_i \neq S'_i$ and $S_j = S'_j$ for all $j \neq i$. A randomized mechanism $\mathcal{M}$ is said to satisfy (ε, δ)-client-level differential privacy if for all such neighboring datasets and for all measurable events $\mathcal{O}$ in the output space,

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

In this paper, we primarily focus on pure $(\varepsilon, 0)$-DP guarantees.

Score Clipping. To ensure bounded sensitivity of the client-level score function required by EM, we apply clipping to the client loss. Specifically, given a clipping threshold $C > 0$, we define the clipped client loss as

$$\tilde{L}_i(\theta) := \text{clip}(L_i(\theta), -C, C),$$

so that $|\tilde{L}_i(\theta)| \leq C$ holds for all clients. This clipping operation is introduced solely to control the sensitivity of the score function for the DP analysis.

3.3 Neighborhood of Worst Client

To analyze how probability mass induced by EM concentrates around the worst-case client, we introduce a notion of neighborhood in the client space, defined in terms of loss gaps.

Loss-Gap–Based Neighborhoods. For any threshold $\delta \geq 0$, we define the δ-neighborhood of the worst client as

$$\mathcal{N}(\delta) := \{i \in \mathcal{I} : \Delta_i \leq \delta\}.$$

By construction, $i^* \in \mathcal{N}(\delta)$ for all $\delta \geq 0$, and the neighborhoods are nested:

$$\delta_1 \leq \delta_2 \Rightarrow \mathcal{N}(\delta_1) \subseteq \mathcal{N}(\delta_2).$$

The set $\mathcal{N}(\delta)$ consists of clients whose losses are within δ of the worst-case loss. This notion of neighborhood captures client-level similarity relative to the worst client. It is agnostic to the underlying data representation and optimization dynamics.

Neighborhood Cardinality and Heterogeneity. Let $N(\delta) := |\mathcal{N}(\delta)|$. The growth of $N(\delta)$ as δ increases provides a coarse characterization of client heterogeneity: small $N(\delta)$ indicates a sharply separated worst client, while large $N(\delta)$ reflects the presence of many near-worst clients. This quantity will play a key role in analyzing probability mass concentration under the exponential mechanism.

Neighborhood Mass. Given the client distribution $p(\varepsilon)$ induced by EM, we define the probability mass assigned to $\mathcal{N}(\delta)$ as:

$$P_\varepsilon(\mathcal{N}(\delta)) := \sum_{i \in \mathcal{N}(\delta)} p_i(\varepsilon). \tag{3}$$

This quantity measures how much probability mass the EM assigns to clients whose losses are within δ of the worst case.

Using the representation $p_i(\varepsilon) \propto \exp(-\varepsilon\Delta_i/(2\tau))$ as in (2), the neighborhood mass depends only on the set of loss gaps $\{\Delta_i\}$ and the privacy parameter ε. Characterizing the behavior of $P_\varepsilon(\mathcal{N}(\delta))$ as a function of (ε, δ) and the gap structure will be the central focus of our analysis in Section 4.

4 Proposed Framework: DP via KL-FDRO

In this section, we first explicitly characterize the probability mass induced by EM (1) with $\mathcal{N}(\delta)$ (Section 4.1). Then, based on this characterization, we re-interpret EM over clients as an in-effect FDRO problem (Section 4.2). Finally, we show that solving FDRO (or equivalently EM) guarantees the DP property (Section 4.3). We also present an illustrative instantiation algorithm in Algorithm 1.

4.1 Neighborhood Mass Concentration

We begin by characterizing how EM (1) induces a continuous spectrum of client emphasis, which concentrates around the worst client (and its neighbors) and ranges from uniform weighting to worst-client dominance, as the privacy parameter ε varies.

Proposition 4.1. *(Neighborhood mass concentration) Let $P_\varepsilon(\mathcal{N}(\delta))$ and $p_i(\varepsilon)$ be defined as in (3) and (1), respectively. The total probability mass assigned to the neighborhood $\mathcal{N}(\delta)$ satisfies*

$$P_\varepsilon(\mathcal{N}(\delta)) = \sum_{i \in \mathcal{N}(\delta)} p_i(\varepsilon) = \frac{\sum_{i:\Delta_i \leq \delta} \exp(-\varepsilon\Delta_i/(2\tau))}{\sum_{j=1}^n \exp(-\varepsilon\Delta_j/(2\tau))}.$$

Moreover, the following bounds hold:

$$\frac{N(\delta)\exp(-\varepsilon\delta/(2\tau))}{N(\delta) + (n - N(\delta))\exp(-\varepsilon\delta/(2\tau))} \leq P_\varepsilon(\mathcal{N}(\delta)) \leq \frac{N(\delta)}{N(\delta)\exp(-\varepsilon\delta/(2\tau)) + (n - N(\delta))\exp(-\varepsilon\Delta_{\max}/(2\tau))},$$

where $\Delta_{\max} := \max_{i \in \mathcal{I}} \Delta_i$.

Proof Sketch. Using the representation in (2), we have $p_i(\varepsilon) \propto \exp(-\varepsilon\Delta_i/(2\tau))$. Summing over all $i \in \mathcal{N}(\delta)$ yields

$$P_\varepsilon(\mathcal{N}(\delta)) = \frac{\sum_{i:\Delta_i \leq \delta} \exp(-\varepsilon\Delta_i/(2\tau))}{\sum_{j=1}^n \exp(-\varepsilon\Delta_j/(2\tau))}.$$

For the lower bound, we use $\Delta_i \leq \delta$ for all $i \in \mathcal{N}(\delta)$ and $\Delta_j \geq \delta$ for all $j \notin \mathcal{N}(\delta)$. For the upper bound, we additionally use $\Delta_j \leq \Delta_{\max}$ for all j. The stated bounds then follow by elementary comparison of the numerator and denominator.

Proposition 4.1 provides a non-asymptotic characterization of the client sampling distribution induced by the exponential mechanism for finite ε, explicitly in terms of the loss-gap structure. These results reveal that EM does not induce a binary transition between average-case and worst-case learning, but instead a graded spectrum of robustness profiles, where probability mass is progressively redistributed from the global population toward increasingly smaller neighborhoods around the worst client.

Proposition 4.1 shows that neighborhood mass concentration under EM is governed by three interacting factors: (i) Privacy parameter ε controls the sharpness of concentration via exponential scaling; (ii) Neighborhood size $N(\delta)$ captures how many near-worst clients compete for probability mass; (iii) Client loss gaps $\Delta_{\max}$ quantify the heterogeneity of the client population.

In particular, as ε increases, probability mass concentrates increasingly on the worst client and its immediate neighborhood, but the rate and extent of this concentration depend critically on the loss-gap structure rather than on the number of clients alone.

We now formalize how the exponential mechanism interpolates between uniform client weighting and worst-client selection as the privacy parameter ε varies.

Corollary 4.2. *(Uniform-to-worst interpolation) Assume the worst client is unique, i.e., $\Delta_{\min} > 0$. The following limits hold:*
1. Uniform regime, $\lim_{\varepsilon \to 0} p_i(\varepsilon) = 1/n, \forall i \in \mathcal{I}$.
2. Worst-client regime, $\lim_{\varepsilon \to \infty} p_{i^*}(\varepsilon) = 1$, $lim_{\varepsilon \to \infty} p_i(\varepsilon) = 0, \forall i \neq i^*$.
Moreover, for any $\varepsilon > 0$, the mapping $\varepsilon \to p_{i^*}(\varepsilon)$ is continuous and strictly increasing.

Proof Sketch. As $\varepsilon \to 0$, we have $\exp(\varepsilon L_i/(2\tau)) \to 1$ for all i, and thus $p_i(\varepsilon) \to 1/n$. As $\varepsilon \to \infty$, since the worst client is unique ($\Delta_{\min} > 0$), all other terms $\exp(-\varepsilon\Delta_i/(2\tau))$ vanish exponentially, implying $p_{i^*}(\varepsilon) \to 1$ and $p_i(\varepsilon) \to 0$ for $i \neq i^*$. Continuity and monotonicity of $p_{i^*}(\varepsilon)$ follow directly from the log-sum-exp form of the normalization.

Quantitative Interpolation Rate. Beyond asymptotic regimes, Proposition 4.1 implies a quantitative interpolation behavior. In particular, for any $\varepsilon > 0$, $p_{i^*}(\varepsilon) = 1/(1 + \sum_{j \neq i^*} \exp(-\varepsilon\Delta_j/(2\tau)))$. This expression highlights that, when the minimum gap $\Delta_{\min}$ is large, concentration on the worst client occurs rapidly as ε increases. When many clients have small gaps, probability mass spreads over a neighborhood rather than collapsing immediately. Corollary 4.2 formalizes the uniform-to-worst spectrum induced by EM in FL: (i) Uniform client averaging $\varepsilon \to 0$ is commonly used in standard FL. Worst-client selection $\varepsilon \to \infty$ is reminiscent of deterministic FDRO. Importantly, intermediate values of ε do not correspond to either

extreme, but instead induce a soft worst-client focus, with probability mass distributed across near-worst clients in a gap-dependent manner. Overall, Section 4.1 shows that EM induces a continuous spectrum of client emphasis, instead of a discrete choice between uniform averaging and worst-case optimization.

4.2 EM as Implicit Client Reweighting

The results in Sections 4.1 suggest that when applied at the client level, EM induces a structured redistribution of importance across clients. In this section, we reinterpret this effect as an implicit client reweighting mechanism, rather than as additive noise or random perturbation.

From Randomization to Reweighting. The distribution $\boldsymbol{p}(\varepsilon)$, depending on θ through $L_i(\theta)$, can be equivalently viewed as defining a set of adaptive mixture weights over clients, where the contribution of each client is modulated by its loss value. Unlike uniform averaging, these weights are loss-dependent and vary continuously with the privacy parameter ε. Under this view, EM does not merely introduce randomness; it reshapes the effective objective by emphasizing high-loss (i.e., worst-case) clients in a controlled, probabilistic manner.

Under this reweighting view, a natural question is how the EM-induced client distribution translates into an effective optimization objective. In particular, as the privacy parameter ε increases, does EM recover a worst-client objective akin to FDRO, or does it induce a fundamentally different form of robustness? The following proposition makes this connection explicit at the objective level.

Proposition 4.3. *(EM-induced soft worst-client objective) Consider the EM-weighted average loss*

$$L_{EM}(\theta;\varepsilon) := \sum_{i=1}^{n} p_i(\varepsilon;\theta)L_i(\theta),\ p_i(\varepsilon;\theta) \propto \exp(\varepsilon L_i(\theta)/(2\tau)).$$

Then, as ε increases, $L_{EM}(\theta;\varepsilon)$ monotonically approaches $\max_i L_i(\theta)$, with the interpolation rate governed by client loss gaps $\{\Delta_i(\theta)\}$.

Proof Sketch. By definition,

$$L_{\mathrm{EM}}(\theta;\varepsilon) = \sum_{i=1}^{n} p_i(\varepsilon;\theta)L_i(\theta) = \nabla_\varepsilon \left(\frac{2\tau}{\varepsilon} \log \sum_{i=1}^{n} \exp(\varepsilon L_i(\theta)/(2\tau)) \right).$$

As ε increases, the log-sum-exp term increasingly concentrates on the largest loss value, implying that $L_{\mathrm{EM}}(\theta;\varepsilon)$ converges monotonically to $\max_i L_i(\theta)$. The convergence rate is governed by the client loss gaps $\{\Delta_i(\theta)\}$.

This result is standard in log-sum-exp analysis and is included here to clarify how EM-induced reweighting realizes a soft form of federated DRO [4].

Comparison with Deterministic Federated DRO. Deterministic federated DRO methods typically focus on minimizing a worst-case or adversarially reweighted client loss, often corresponding to

$$\max_{i\in I} L_i \text{ or } \sup_{q\in Q} \sum_{i=1}^{n} q_i L_i,$$

for some uncertainty set $Q \subseteq \mathrm{Simplex}_n$, e.g., KL-ball around uniform. Such formulations focus exclusively on the extreme clients permitted by Q, and the resulting objective depends only on the maximal attainable risk, instead of on the full profile of relative loss gaps among clients.

In contrast, the EM induces a soft and gap-sensitive alternative that admits an exact variational characterization in terms of a KL-regularized DRO problem. Specifically, for fixed model parameter θ, the log-sum-exp objective admits the following KL-regularized DRO representation:

$$\frac{2\tau}{\varepsilon} \log\Big(\frac{1}{n}\sum_{i=1}^{n} \exp(\varepsilon L_i/(2\tau))\Big) = \sup_{q\in \mathrm{Simplex}_n} \sum_{i=1}^{n} q_i L_i - \frac{2\tau}{\varepsilon}\mathrm{KL}(q||u), \tag{4}$$

where u denotes the uniform distribution over clients. Due to the strict convexity of the KL regularizer, the optimal solution q^* to this problem coincides exactly with the exponential-mechanism sampling distribution $\boldsymbol{p}(\varepsilon)$. This variational representation is well known in DRO with KL-divergence uncertainty sets [2, 4]. It highlights a fundamental distinction between deterministic federated DRO and the exponential-mechanism-based approach: while deterministic DRO emphasizes worst-case clients through hard uncertainty sets, the exponential mechanism realizes a regularized DRO that smoothly interpolates between average-risk minimization and worst-case robustness, with the privacy parameter ε controlling the strength of this bias via a KL divergence penalty.

Bias–Privacy Tradeoff. From the reweighting perspective, the privacy parameter ε admits a natural interpretation as a bias control parameter. For small ε, weights are nearly uniform, leading to minimal bias toward any specific client. For large ε, weights concentrate on the worst client and its immediate neighborhood, inducing a strong robustness bias.

Importantly, this bias is structured, rather than arbitrary: Proposition 4.1 shows that it depends explicitly on client loss gaps rather than on the number of clients alone. This observation highlights a previously underexplored aspect of privacy mechanisms, i.e., even when introduced for privacy protection, they can systematically reshape the optimization landscape.

Relation to Noise-Based DP Mechanisms. Traditional differentially private FL algorithms enforce privacy by injecting noise at each iteration or aggregation step. Such noise is often treated as an optimization artifact whose primary effect is to degrade utility. In contrast, EM studied in this paper induces a structured and loss-dependent sampling bias at the client level, which corresponds to a FDRO-type objective in terms of effective client weighting and gradient aggregation. Instead of perturbing gradients or model updates, it alters which clients are emphasized in the learning process. This distinction suggests that privacy mechanisms may influence learning behavior not only through variance inflation, but also through implicit prioritization of certain clients.

An Illustrative Instantiation of the EM Algorithm. While this paper primarily focuses on the theoretical characterization of EM-induced robustness in FL, it is instructive to see how the above perspective can be instantiated in a concrete federated optimization procedure. To this end, Algorithm 1 presents a simple example of an EM-based FDRO framework, where EM is used to induce client-level reweighting during training. We emphasize that Algorithm 1 is not meant to be an optimized or complete algorithmic solution,

Algorithm 1 A minimal instantiation of EM-induced client reweighting under the FDRO perspective

Require: model θ_0, privacy budget ε, learning rate η_t, number of selected clients m, sensitivity bound τ.

1: **for** $t = 0, ..., T-1$ **do**
2: Broadcast model: server sends θ_t to all clients.
3: Each client i computes its client-level loss

$$L_i(\theta_t) = \mathbb{E}_{(x,y)\sim\mathcal{D}_i}[\ell(\theta_t; x, y)],$$

or its empirical estimate $\tilde{L}_i(\theta_t)$ on a batch $B_{t,i}$
4: Server samples a subset of clients C_t of size m according to the EM probability $p_i^t(\varepsilon; \theta_t)$ as in (1)
5: Each selected client $i \in C_t$ computes a local update $g_{i,t}(\theta_t)$, e.g., a stochastic gradient
6: Server updates the model:

$$\theta_{t+1} = \theta_t - \eta_t \sum_{i\in C_t} g_{i,t}(\theta_t).$$

7: **end for**
8: **return** θ_T

but rather a minimal instantiation that illustrates how EM-induced reweighting can be integrated into standard FL pipelines. While Algorithm 1 abstracts away many implementation details of local training, the clipping operation is introduced solely to bound the sensitivity of the client-level score for the DP analysis in Section 4.3, and is not intrinsic to the FDRO interpretation itself.

4.3 Client-Level DP via FDRO

To study the client-level DP property, we consider a client-level adjacency relation standard in federated learning: two federated datasets $S = (S_1, ..., S_n)$ and $S' = (S'_1, ..., S'_n)$ are said to be neighboring if they differ in exactly one client dataset, i.e., there exists an index i such that $S_i \neq S'_i$ and $S_j = S'_j$ for all $j \neq i$. Our goal is to ensure differential privacy with respect to this adjacency notion. This means that the observable behavior of the algorithm, i.e., most notably the client selection and weighting induced by the exponential mechanism, does not change significantly when a single client's dataset is modified or replaced.

We analyze the client-level DP of the EM-based client selection step used in Algorithm 1, which constitutes the core source of randomness and privacy protection in this instantiation.

Proposition 4.4. *(Client-level DP via EM-FDRO) Consider the client-level score function $s_i(S) = \tilde{L}_i(\theta_t)$, where $\tilde{L}_i$ is a clipped client loss satisfying $|\tilde{L}_i| \leq C$. Under the client-level adjacency relation defined above, assume the score sensitivity is upper bounded by $\tau \leq \tau_0$. Then, in each round t, the EM that samples clients according to $p_i^t \propto \exp(\varepsilon s_i(S)/(2\tau_0))$, as in (1), satisfies $(\varepsilon, 0)$-DP w.r.t. the client datasets S.*

Moreover, over T rounds, basic composition yields $(T\varepsilon, 0)$-DP for the sequence of client selections.

Proof Sketch. Under the assumed clipping, the client-level score function $s_i(S) = \tilde{L}_i(\theta_t)$ has sensitivity at most τ_0 with respect to the client-level adjacency relation. By the standard guarantee of the exponential mechanism [6, 18], sampling clients according to $p_i \propto \exp(\varepsilon s_i/(2\tau_0))$ satisfies $(\varepsilon, 0)$-DP for each round. The multi-round guarantee follows directly from basic composition.

In practice, the exact client loss $L_i(\theta_t)$ may not be available due to stochastic estimation, mini-batching or communication constraints [16]. This leads to an approximate implementation of the EM, whose privacy implications we now characterize in the following Lemma 4.5 and Corollary 4.6.

In practice, the exact losses may be unavailable, yielding an approximate EM implementation. We next quantify how loss approximation error degrades the per-round DP guarantee. Specifically, let the implemented loss of $L_i(\theta_t)$ be $\tilde{L}_i(\theta_t) = L_i(\theta_t) + \epsilon_i^t$, for $|\epsilon_i^t| \leq \nu, \forall i$, where ϵ_i^t is the approximation error and ν is the upper bound of such approximation error.

Lemma 4.5. *(Approximate EM implies approximate sampling distribution) Let $\boldsymbol{p}$ be the ideal EM distribution using L_i, and $\tilde{\boldsymbol{p}}$ be the implemented EM distribution using $\tilde{L}_i$. Then for any i, the following inequalities hold:*

$$\exp(-\gamma_t) \leq \frac{\tilde{p}_i}{p_i} \leq \exp(\gamma_t), \quad \text{with } \gamma_t := \frac{\varepsilon\nu}{2\tau}.$$

Equivalently, $\tilde{p}$ is within a multiplicative factor $\exp(\pm\gamma_t)$ of the correct distribution $\boldsymbol{p}$.

Proof Sketch. The approximation error perturbs the exponent of each term in the EM distribution by at most $\varepsilon\nu/(2\tau)$. Taking ratios between the ideal and implemented distributions yields a multiplicative distortion bounded by $\exp(\pm\gamma_t)$, where $\gamma_t = \varepsilon\nu/(2\tau)$.

Corollary 4.6. *(Approximate EM implies degraded DP) If the correct per-round EM selection is ε-DP, and the implemented selection distribution has multiplicative distortion $\exp(\pm\gamma_t)$, then the implemented selection is $(\varepsilon + 2\gamma_t, 0)$-DP.*

Proof Sketch. A multiplicative distortion of $\exp(\pm\gamma_t)$ in the output distribution corresponds to an additive increase of $2\gamma_t$ in the privacy parameter under pure DP. The result follows by standard arguments on approximate implementations of differentially private mechanisms.

This result shows that implementation errors in client scoring translate linearly into privacy degradation. Importantly, the degradation depends on the score approximation error, instead of on the number of clients. This highlights the robustness of EM-based DP guarantees under practical approximations.

5 Experiments

The goal of this experiment is to empirically validate *whether EM–inspired client sampling, which is originally designed to guarantee DP, can serve as a controllable robustness knob in FL at the same time.* In particular, we aim to study how the privacy parameter ε modulates the concentration of client selection, and how such bias toward higher-loss (hard-to-fit) clients affects both global utility and tail-client performance under heterogeneous data distributions.

Therefore, the experiment is designed to examine the trade-offs induced by biased client sampling (from EM), including (i) changes in optimization dynamics, (ii) robustness to data heterogeneity and (iii) performance on underrepresented clients. We may not necessarily examine and observe any unconditional accuracy improvements.

5.1 Experiment Setup

Model and Dataset. We consider a standard federated image classification benchmark using CIFAR-10 [14] with a ResNet-18 model [10]. The convolutional stem is modified to match CIFAR resolution (kernel size 3×3, stride 1, no max pooling), while the rest of the backbone architecture follows the standard ResNet-18 design.

The federated system consists of 25 clients, each holding a non-IID subset of the training data generated via a Dirichlet partition with concentration parameter $\alpha = 0.3$, inducing moderate label skew across clients. Each client holds about 2,000 samples with comparable local dataset sizes. Statistical heterogeneity is introduced via Dirichlet label-skew partitioning ($\alpha = 0.3$), and each client further reserves 10% of its local data as a validation set. At each communication round, 20 clients are selected to participate in training, each performing 2 local epochs of SGD with batch size 256, learning rate 0.05, momentum 0.9, and weight decay 5×10^{-4}. The global model is evaluated on the centralized CIFAR-10 test set after every round.

EM Sampling over Clients. Client selection is performed using an EM–style distribution:

$$p_i \propto \exp(\varepsilon L_i^{\text{val}}/(2\tau)),$$

where L_i^{val} denotes the most recently observed validation loss of client i (analogous to L_i defined in Section 3.1), τ is the sensitivity parameter of EM, and ε controls the degree of sampling bias. In our experiment, we fix $\tau = 1$ for simplicity, while we compare $\varepsilon = \{0, 2, 4\}$, where $\varepsilon = 0$ corresponds to the uniform client sampling and reduces to FedAvg [16] baseline.

Evaluation Metrics. We report the following metrics to capture both utility and robustness:

- Global test accuracy on the centralized CIFAR-10 test set.
- Worst client accuracy, to measure performance on hardest or most underrepresented client.
- Tail client accuracy, defined as the 10th percentile of per-client validation accuracy, to measure performance on harder or underrepresented clients.
- Sampling concentration, measured by the maximum selection probability $p_{\max}$, and the top-k probability mass, to quantify how strongly client selection departs from uniformity as ε increases.

To mitigate stochasticity induced by non-IID data partitioning and stochastic optimization, we emphasize trajectory-level behavior rather than single-round or peak performance. Accordingly, all metrics shown in Figure 3 are computed by averaging per-round values over the entire training trajectory and then aggregated across 4 random seeds, while we provide results on a seed in Figure 2. This protocol yields a stable and reproducible comparison of client-level performance under different sampling mechanisms.

5.2 Experiment Results

5.2.1 Learning Dynamics on a Random Run. Figure 2 shows the four evaluation metrics with a random seed using a moving-average aggregation over a window size of 5 rounds (totally 30 rounds). Our observations are as follows.

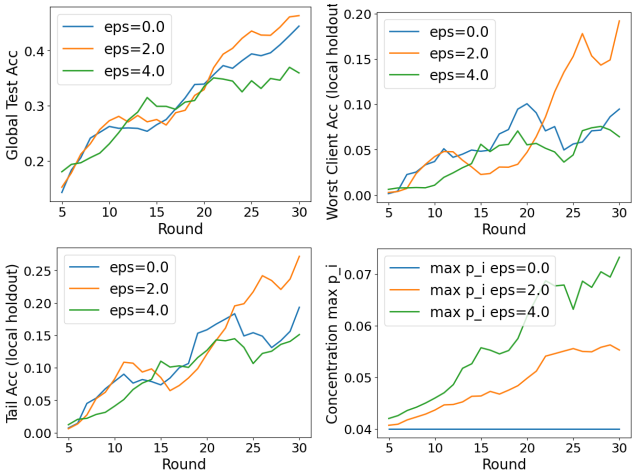

Figure 2: Results on a run during the entire learning dynamics (30 rounds). Top left: global (unconditional) test accuracy. Top right: worst client accuracy. Bottom left: tail client accuracy. Bottom right: maximum selection probability $p_{\max}$ of EM. We draw three observations: (i) appropriate $\varepsilon = 2$ achieves the best global test accuracy, though there are not very significant gaps among all three options of ε. (ii) worst and tail client test accuracy both demonstrates that $\varepsilon = 2$ improves the robustness with a significant margin compared to other two baselines. (iii) the $p_{\max}$ verifies that a larger ε makes the sampling mass probability more concentrated on the worst client.

(1). Figure 2 (top-left) shows that EM-based client sampling preserves stable global convergence across all $\varepsilon = \{0, 2, 4\}$ values. Notably, intermediate bias ($\varepsilon = 2$) achieves comparable or slightly higher global accuracy than uniform sampling ($\epsilon = 0$), while excessive bias ($\epsilon = 4$) does not yield further gains. This behavior is consistent with our theoretical analysis in Section 4.1 and 4.2, which predicts that intermediate ε values induce a soft worst-client emphasis rather than collapsing onto a single client. Such soft reweighting can improve robustness without significantly sacrificing global generalization.

(2). Figure 2 (top-right and bottom-left) demonstrates substantial improvements in worst-client and tail-client accuracy under moderate EM bias ($\varepsilon = 2$). Compared to uniform sampling, EM increases the participation frequency of high-loss clients and their near-worst neighbors, which aligns with the neighborhood mass concentration behavior characterized in Proposition 4.1. Importantly, the performance gains are not limited to a single extreme client, but extend to a broader set of near-worst clients, supporting our claim that EM induces a gap-sensitive robustness profile rather than a hard worst-client focus.

(3). Figure 2 (bottom-right) reports the maximum client selection probability $p_{\max}$, which serves as an empirical proxy for sampling concentration. As ε increases, ε grows smoothly from approximately $1/n$ toward larger values, confirming the continuous interpolation behavior predicted by Corollary 4.2. Crucially, the increase in $p_{\max}$

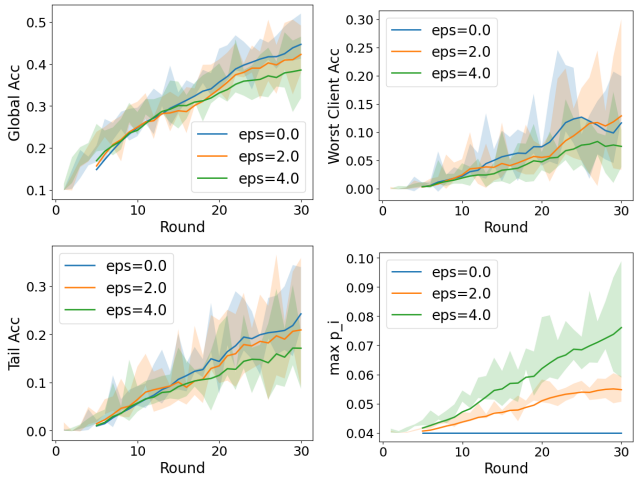

Figure 3: Results averaged over 4 runs. Top-left: global test accuracy. Top-right: worst-client accuracy. Bottom-left: tail-client accuracy (10th percentile). Bottom-right: maximum selection probability $p_{\max}$ of EM. For each seed, per-round trajectories are first smoothed using a moving average with window size 5, and shaded regions indicate one standard deviation across seeds. As the privacy parameter ε increases, client sampling becomes progressively more concentrated (larger $p_{\max}$). Moderate bias ($\varepsilon = 2$) consistently improves worst-client performance without harming global accuracy, while excessive bias ($\varepsilon = 4$) leads to over-concentration and degraded robustness. Notably, tail-client performance does not monotonically improve with ε, highlighting a trade-off between extreme worst-client emphasis and broader tail robustness.

is gradual rather than abrupt, indicating that EM does not immediately collapse onto a single worst client, but instead redistributes probability mass over a neighborhood of high-loss clients.

5.2.2 Learning Dynamics over Multiple Runs. To assess the robustness of the observed trends, we aggregate results over 4 independent random seeds. Figure 3 reports the mean trajectories with standard deviation bands across seeds for all evaluation metrics.

(1). Figure 3 (top-left) shows that EM-based client sampling preserves stable global convergence across all $\varepsilon \in \{0, 2, 4\}$. Similar to the single-run behavior in Figure 2, intermediate bias ($\varepsilon = 2$) achieves comparable or slightly higher global test accuracy than uniform sampling ($\varepsilon = 0$), while excessive bias ($\varepsilon = 4$) does not yield further gains. This confirms that EM-induced reweighting does not compromise overall utility under moderate bias.

(2). Figure 3 (top-right) demonstrates that moderate EM bias ($\varepsilon = 2$) consistently achieves the highest worst-client accuracy across seeds, particularly in later training rounds. This indicates that mildly prioritizing high-loss clients can robustly improve performance on the single most challenging client, aligning with the neighborhood mass concentration behavior characterized in Proposition 4.1. In contrast, Figure 3 (bottom-left) shows that tail-client accuracy (10th percentile) exhibits a different trend. Uniform sampling ($\varepsilon = 0$) achieves the best tail performance on average, while $\varepsilon = 2$ remains competitive but does not uniformly outperform FedAvg. Aggressive bias ($\varepsilon = 4$) leads to degraded tail performance. These results suggest that EM primarily benefits the extreme worst client, while excessive concentration may reduce coverage over a broader set of near-worst clients.

(3). Figure 3 (bottom-right) reports the maximum client selection probability $p_{\max}$ as an empirical proxy for sampling concentration. As ε increases, $p_{\max}$ grows smoothly from approximately $1/n$ toward larger values, confirming the continuous interpolation behavior predicted by Corollary 4.2. Importantly, the increase in $p_{\max}$ is gradual rather than abrupt, indicating that EM redistributes probability mass over a neighborhood of high-loss clients instead of immediately collapsing onto a single worst client.

Overall, the multi-run results in Figure 3 reinforce our central finding: EM induces a controllable spectrum of robustness behaviors governed by the privacy parameter ε. While moderate bias can consistently improve worst-client performance without harming global accuracy, stronger bias leads to over-concentration and diminished tail robustness, highlighting an inherent trade-off between extreme worst-client emphasis and broader client-level fairness.

6 Conclusion

In this paper, we established a principled connection between client-level DP and FDRO through EM. By analyzing the concentration behavior of EM over heterogeneous clients, we showed that privacy-induced randomization yields a continuous and gap-dependent spectrum of robustness profiles controlled by the privacy parameter ε. This perspective clarifies how privacy constraints implicitly reshape client weighting and robustness in FL. Empirical results on CIFAR-10 federated benchmarks illustrate this trade-off, showing that moderate privacy bias can improve worst-client performance without harming global utility, while excessive concentration degrades broader tail robustness. Our findings provide a unified view of privacy and robustness in FL and suggest new directions for designing DP-aware FL optimization methods beyond noise-based perturbations.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 308–318.

[2] Aharon Ben-Tal, Dick Den Hertog, Anja De Waegenaere, Bertrand Melenberg, and Gijs Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* 59, 2 (2013), 341–357.

[3] Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. 2020. Distributionally robust federated averaging. *Advances in neural information processing systems* 33 (2020), 15111–15122.

[4] John C Duchi and Hongseok Namkoong. 2021. Learning models with uniform performance via distributionally robust optimization. *The Annals of Statistics* 49, 3 (2021), 1378–1406.

[5] Cynthia Dwork and Jing Lei. 2009. Differential privacy and robust statistics. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*. 371–380.

[6] Cynthia Dwork and Aaron Roth. 2014. *The algorithmic foundations of differential privacy*. Vol. 9. Foundations and Trends® in Theoretical Computer Science. 211–407 pages.

[7] Robin C Geyer, Tassilo Klein, and Moin Nabi. 2017. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557* (2017).

[8] Zhishuai Guo and Tianbao Yang. 2024. Communication-efficient federated group distributionally robust optimization. *Advances in Neural Information Processing Systems* 37 (2024), 23040–23077.

[9] Andrew Hard, Kanishka Rao, Rajiv Mathews, Swaroop Ramaswamy, Françoise Beaufays, Sean Augenstein, Hubert Eichner, Chloé Kiddon, and Daniel Ramage. 2018. Federated learning for mobile keyboard prediction. *arXiv preprint arXiv:1811.03604* (2018).

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[11] Chaouki Ben Issaid, Anis Elgabli, and Mehdi Bennis. 2022. DR-DSGD: A distributionally robust decentralized learning algorithm over graphs. *arXiv preprint arXiv:2208.13810* (2022).

[12] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning* 14, 1–2 (2021), 1–210.

[13] Jakub Konečnỳ, H Brendan McMahan, Daniel Ramage, and Peter Richtárik. 2016. Federated optimization: Distributed machine learning for on-device intelligence. *arXiv preprint arXiv:1610.02527* (2016).

[14] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).

[15] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems* 2 (2020), 429–450.

[16] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[17] H Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. 2018. Learning Differentially Private Recurrent Language Models. In *International Conference on Learning Representations*.

[18] Frank McSherry and Kunal Talwar. 2007. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 94–103.

[19] Kentaro Minami, HItomi Arai, Issei Sato, and Hiroshi Nakagawa. 2016. Differential privacy without sensitivity. *Advances in Neural Information Processing Systems* 29 (2016).

[20] Mehryar Mohri, Gary Sivek, and Ananda Theertha Suresh. 2019. Agnostic federated learning. *arXiv preprint arXiv:1902.00146* (2019).

[21] Hongseok Namkoong and John C Duchi. 2017. Variance-based regularization with convex objectives. In *Advances in neural information processing systems*. 2971–2980.

[22] Sarthak Pati, Ujjwal Baid, Brandon Edwards, Micah Sheller, Shih-Han Wang, G Anthony Reina, Patrick Foley, Alexey Gruzdev, Deepthi Karkada, Christos Davatzikos, et al. 2022. Federated learning enables big data for rare cancer boundary detection. *Nature communications* 13, 1 (2022), 7346.

[23] Anastasia Pustozerova, Jan Baumbach, and Rudolf Mayer. 2023. Differentially private federated learning: Privacy and utility analysis of output perturbation and dp-sgd. In *2023 IEEE International Conference on Big Data (BigData)*. IEEE, 5549–5558.

[24] Hamed Rahimian and Sanjay Mehrotra. 2019. Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659* (2019).

[25] Minseok Ryu and Kibaek Kim. 2022. Differentially private federated learning via inexact ADMM with multiple local updates. *arXiv preprint arXiv:2202.09409* (2022).

[26] Aras Selvi, Huikang Liu, and Wolfram Wiesemann. 2025. Differential privacy via distributionally robust optimization. *Operations Research* (2025).

[27] Stacey Truex, Ling Liu, Ka-Ho Chow, Mehmet Emre Gursoy, and Wenqi Wei. 2020. LDP-Fed: Federated learning with local differential privacy. In *Proceedings of the third ACM international workshop on edge systems, analytics and networking*. 61–66.

[28] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H Yang, Farhad Farokhi, Shi Jin, Tony QS Quek, and H Vincent Poor. 2020. Federated learning with differential privacy: Algorithms and performance analysis. *IEEE transactions on information forensics and security* 15 (2020), 3454–3469.

[29] Xi Wu, Fengan Li, Arun Kumar, Kamalika Chaudhuri, Somesh Jha, and Jeffrey Naughton. 2017. Bolt-on differential privacy for scalable stochastic gradient descent-based analytics. In *Proceedings of the 2017 ACM international conference on management of data*. 1307–1322.