VALUE BEFORE DISCLOSURE: DATA VALUATION UNDER LOCAL DIFFERENTIAL PRIVACY

Anonymous authors
Paper under double-blind review

ABSTRACT

Data privacy is a critical concern in data valuation. Assessing a record's task-dependent value before purchase often requires access to potentially sensitive information. We propose PIVOTS, a local differential privacy (LDP) framework that lets providers estimate their contributions without first disclosing their raw records. The buyer uses one locally randomized report from each provider and validation data to construct a broadcast valuation function that each provider evaluates on its retained record to estimate its value. Theoretically, we establish mean-squared-error upper bounds and matching minimax lower bounds when validation data are sufficiently abundant. We also consider settings in which some non-sensitive features remain public, developing a corresponding valuation method and providing theoretical guarantees that quantify the resulting gain in valuation accuracy brought by partial feature disclosure. Experiments on synthetic and real-world data demonstrate accurate valuation under privacy constraints and substantial improvements in downstream task performance.

1 INTRODUCTION

Machine learning systems draw on data from many sources, yet individual records can contribute very differently to a given task. A useful record can improve a target model, while a redundant or corrupted one can add little or even harm it (Ghorbani & Zou, 2019; Jia et al., 2019b). Data valuation measures these task-dependent contributions to guide data acquisition and compensation to data providers (Agarwal et al., 2019). In a standard data-pricing workflow, the buyer first estimates the contribution of each seller's record to its task, taking other candidate records into account, then offers a price. Such estimation requires the buyer to acquire sufficient information from the seller.

Direct disclosure of a sensitive record allows the buyer to inspect its contents and assess its likely usefulness before making an offer, exposing the seller's private information and potentially weakening its bargaining position (Figure 1A). However, without information about the candidate pool, the buyer cannot account for redundancy among records and may therefore misprice them (Figure 1B). To address the challenge of valuing data while limiting disclosure, prior work has used established privacy-preserving techniques, including differential privacy (Wang et al., 2023; Sim et al., 2023) and cryptographic methods (Zheng et al., 2023; Peng et al., 2025). However, these methods either rely on a trusted central curator or require the deployment of encryption protocols. For buyers purchasing individual records from a large, dispersed user population, these requirements may be impractical. This motivates our research question: *Can a seller estimate the value of its own record before disclosing it to any other party?*

In this work, we answer this question by proposing PIVOTS (**P**rovider-side pr**IV**ate valuati**O**n with **T**knn-**S**hapley), a data valuation framework that enables local data valuation with guaranteed privacy protection. The buyer announces the task and the valuation rule. Each provider then sends one locally randomized report that satisfy the requirement of local differential privacy (LDP). The buyer aggregates the reports into a summary of the candidate pool, which, together with validation data, yields a function that maps each provider's retained record to an estimated value and is then broadcast to the providers. Each provider applies the function to its retained record to estimate the record's value locally, then decides whether to transact (Figure 1C).

We summarize our contribution as follows:

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

- We formalize locally private data valuation as a provider-side estimation problem and propose a framework that lets providers estimate their own value before raw disclosure. We also discuss settings in which some non-sensitive features remain public (Sections 2–3).
- We propose PIVOTS, a non-interactive locally differentially private valuation method that requires only one randomized report per provider, accommodates different settings of public and private features, and enables providers to obtain reliable value estimates locally without disclosing their raw records (Section 3).
- Under mild distributional assumptions, we derive upper bounds on the mean squared error of our valuation. Matching minimax lower bounds show that the resulting rates are optimal for the specified reporting protocols (Section 4).
- We conduct experiments on synthetic and real datasets to examine theoretical error trends and evaluate value recovery, corruption detection, and downstream performance after data selection. Our method significantly outperforms naive baselines (Section 5).



(A) Raw-data disclosure. (B) Blind offer. (C) PIVOTS framework.

Figure 1: Examples of data valuation framework. In (A), raw information disclosure exposes provider records before a price is known. In (B), a blind server sends the same offer without information of the candidate data pool, where all three providers hold similar records and should therefore receive lower prices because of redundancy. In PIVOTS (C), providers send locally randomized information, the server aggregates and returns a pricing rule, and each provider uses its retained record to estimate its own value locally.

2 Preliminaries

2.1 Data valuation

Let $D = \{z_i\}_{i=1}^N$ denote a fixed candidate dataset held by N providers, where provider i holds $z_i = (x_i, y_i) \in [0,1]^d \times \{0,1\}$. The buyer holds a validation set $D^{(\mathrm{val})} = \{z_j^{(\mathrm{val})}\}_{j=1}^M$ of i.i.d. samples from a distribution Q, independently of D. Data valuation aims to assign each record z_i a value reflecting its contribution to the performance of a buyer's task, as evaluated under Q.

Shapley-based valuation is a well-established framework for assigning values to training records (Ghorbani & Zou, 2019; Jia et al., 2019b). It measures each record's contribution by averaging its marginal effect on a specified learning utility across training subsets. This averaging captures how a record's value depends on the other available records, including redundancy and complementarity, rather than evaluating it in isolation. Subsequent work on KNN-Shapley and TKNN-Shapley shows that Shapley values for nearest-neighbor utilities can be computed or approximated efficiently (Jia et al., 2019a; Wang et al., 2023). This tractability makes Shapley-based valuation a practical starting point for the locally private strategy developed below.

Shapley value. Let D_{-i} be D with provider i's indexed record removed, and let $|A|$ denote the number of records in A. For any utility function U defined on subsets of D, the Shapley value of provider i's record is

$$\phi_i(D_{-i}; U) = \frac{1}{N} \sum_{k=1}^{N} \binom{N-1}{k-1}^{-1} \sum_{A \subseteq D_{-i}, |A|=k-1} \big[U(A \cup \{z_i\}) - U(A)\big]. \tag{1}$$

In (1), $\binom{N-1}{k-1}^{-1}$ averages the marginal gain $U(A \cup \{z_i\}) - U(A)$ over all subsets $A \subseteq D_{-i}$ of size $k-1$. For any evaluation point v, let $U_v(A)$ be the utility of a training subset A evaluated at v. Write $\phi_i(v) := \phi_i(D_{-i}; U_v)$. The noise-free validation-sample value is $\widehat{\Phi}_i := M^{-1} \sum_{j=1}^{M} \phi_i(z_j^{(\mathrm{val})})$. It uses the unrandomized candidate records and is an analytical benchmark rather than a value available to the buyer. Its population counterpart, conditional on the realized D, is

$$\Phi_i^\star = \int \phi_i(v)\, Q(dv) = \mathbb{E}_{v \sim Q}\big[\phi_i(v)\big]. \tag{2}$$

Equation (2) averages over the buyer's evaluation distribution Q while holding D fixed. It measures a candidate record's expected contribution to future evaluations with other candidates available

KNN and TKNN-Shapley. Direct evaluation of (1) is computationally prohibitive for large N, as it sums over all 2^{N-1} subsets of D_{-i}. For k-nearest-neighbor utilities, Jia et al. (2019a) and Wang & Jia (2023b) provide efficient algorithms for exact and approximate KNN-Shapley computation. The exact algorithms sort training records by distance to each validation point and compute their values recursively.

We use TKNN-Shapley as the basis of our method because its threshold-based formulation is amenable to locally private estimation. For each validation point, the TKNN utility evaluates a training subset using only its records within distance τ of that point (Wang et al., 2023). Specifically, let $z^{(\mathrm{val})} = (x^{(\mathrm{val})}, y^{(\mathrm{val})})$ be a validation point and fix a threshold $\tau > 0$. Let $\mathrm{NB}_{x^{(\mathrm{val})},\tau}(A) = \{(x, y) \in A : \|x - x^{(\mathrm{val})}\|_2 \le \tau\}$ denote the records of A within distance τ of $x^{(\mathrm{val})}$. The pointwise TKNN utility is

$$v_{z^{(\mathrm{val})}}^{\mathrm{TKNN}}(A) = \begin{cases} \frac{1}{2}, & |\mathrm{NB}_{x^{(\mathrm{val})},\tau}(A)| = 0, \\ \dfrac{1}{|\mathrm{NB}_{x^{(\mathrm{val})},\tau}(A)|} \displaystyle\sum_{(x,y) \in \mathrm{NB}_{x^{(\mathrm{val})},\tau}(A)} \mathbf{1}[y = y^{(\mathrm{val})}], & |\mathrm{NB}_{x^{(\mathrm{val})},\tau}(A)| > 0. \end{cases} \tag{3}$$

Substituting this utility for U_v in (1) yields the pointwise TKNN-Shapley value $\phi_i(v)$. Given z_i, $\phi_i(v)$ depends on the other records only through the neighbor count and the matching-label count. Thus, pointwise value estimation reduces to computing these two counts. This count-based formulation will later allow us to cast pointwise value estimation under LDP as the well-studied problem of private count estimation (Kairouz et al., 2016a; Wang et al., 2017; Bassily, 2019).

2.2 Local Differential Privacy

We now define local differential privacy, which we later use to formally establish the privacy guarantees of our reporting mechanisms.

Definition 2.1 (α-local differential privacy (α-LDP))**.** Let $\mathcal{Z} = [0,1]^d \times \{0,1\}$. For $\alpha > 0$, a randomized mechanism $\mathcal{M}$ is α-LDP if, for every $z, z' \in \mathcal{Z}$ and every measurable output event E,

$$\Pr\{\mathcal{M}(z) \in E\} \le e^{\alpha} \Pr\{\mathcal{M}(z') \in E\}. \tag{4}$$

For any z, z' and report event E with positive probability, (4) and its reverse bound the likelihood ratio for E between $e^{-\alpha}$ and e^{α}. Thus, any observed report changes the relative odds of z and z' by at most e^{α}, giving the provider plausible deniability. In this paper, we only consider non-interactive LDP: each provider sends one report based only on its own z_i, without using other providers' messages. Sequentially interactive and fully interactive protocols are outside our scope.

The LDP guarantee above protects every component of a record, which can be stronger than needed when some features are public or not sensitive (Krichene et al., 2024; Chua et al., 2024; Ma et al., 2026). To accommodate such settings, we also study *semi-LDP*, which protects the label and designated private features while allowing the remaining features to be released, and investigate whether this selective protection improves data valuation accuracy.

Definition 2.2 (α-semi-LDP)**.** Suppose there is a publicly known split of the feature coordinates with $s \in \{0, \dots, d\}$ private coordinates and $d - s$ public coordinates. Write $z = (x^{\mathrm{priv}}, x^{\mathrm{pub}}, y) \in \mathcal{Z}$, where $x^{\mathrm{priv}} \in [0,1]^s$ and $x^{\mathrm{pub}} \in [0,1]^{d-s}$. A mechanism $\mathcal{M}$ is α-semi-LDP if for every $z = (x^{\mathrm{priv}}, x^{\mathrm{pub}}, y), z' = (x'^{\mathrm{priv}}, x^{\mathrm{pub}}, y') \in \mathcal{Z}$ and every measurable output event E,

$$\Pr\{\mathcal{M}(z) \in E\} \le e^{\alpha} \Pr\{\mathcal{M}(z') \in E\}. \tag{5}$$

The privacy guarantee compares report distributions as (x^{priv}, y) varies while x^{pub} is held fixed and treated as public. This definition includes Label LDP (Xu et al., 2023; Zhao et al., 2025a) when $s = 0$ and Full LDP when $s = d$ as special cases.

3 METHODOLOGY OF PIVOTS

In this section, we present the methodology of PIVOTS. Section 3.1 expresses pointwise TKNN-Shapley through two counts and bounds its sensitivity to them. Section 3.2 constructs histogram counts and estimates them from one private report per provider. Section 3.3 stabilizes these estimates and forms the final valuation. Section 3.4 discusses how to protect the buyer's validation data.

3.1 A SAMPLE-LEVEL LOCAL-COUNT REPRESENTATION

The exact, non-private sample value is $\widehat{\Phi}_i = M^{-1} \sum_{t=1}^{M} \phi_i(z_t^{(\mathrm{val})})$. To compute this value, we first express the contribution at one validation point through local counts. Fix $v = (u, r)$, with feature vector u and label $r \in \{0, 1\}$, and set $I_k(u) = \mathbf{1}[\|x_k - u\|_2 \le \tau]$ and $R_k(u, r) = I_k(u)\mathbf{1}[y_k = r]$. The leave-one-out neighborhood counts are $C_{-i}(u) = \sum_{k \neq i} I_k(u)$, the number of other neighbors, and $S_{-i}(u, r) = \sum_{k \neq i} R_k(u, r)$, the number of those neighbors whose labels match r.

To compute $\phi_i(v)$, consider i's marginal gain in a uniformly random order of the providers. Non-neighbors do not affect the TKNN utility, so $\phi_i(v) = 0$ if $I_i(u) = 0$. If $I_i(u) = 1$, let $\pi = (\pi_1, \ldots, \pi_{C_{-i}(u)})$ list the other neighbors in their order, and let m be the number appearing before i. The $C_{-i}(u) + 1$ possible values of m are equally likely. When $m = 0$, adding i changes the utility from $1/2$ to $R_i(u, r)$, giving a marginal gain of $R_i(u, r) - 1/2$. When $m \ge 1$, its marginal gain is

$$\frac{R_i(u, r) + \sum_{\ell=1}^{m} R_{\pi_\ell}(u, r)}{m + 1} - \frac{\sum_{\ell=1}^{m} R_{\pi_\ell}(u, r)}{m} = \frac{R_i(u, r) - m^{-1} \sum_{\ell=1}^{m} R_{\pi_\ell}(u, r)}{m + 1}.$$

For $m \ge 1$, the first m neighbors form a uniformly chosen subset of the $C_{-i}(u)$ other neighbors, so their expected matching-label fraction is $S_{-i}(u, r)/C_{-i}(u)$. Averaging over m gives the exact TKNN-Shapley value (Wang et al., 2023, Theorem 13). The detailed derivation is in Appendix B. Specifically, $\phi_i(v) = G(I_i(u), R_i(u, r), C_{-i}(u), S_{-i}(u, r))$, where, for feasible inputs $I \in \{0, 1\}$, $R \in \{0, I\}$, and $0 \le S \le C$,

$$G(I, R, C, S) = \begin{cases} 0, & I = 0, \\ R - 1/2, & I = 1,\ C = 0, \\ \dfrac{1}{C+1}\left[R - \dfrac{1}{2} + \displaystyle\sum_{m=1}^{C} \frac{R - S/C}{m+1}\right], & I = 1,\ C \ge 1. \end{cases} \tag{6}$$

Equation (6) makes valuation a plug-in problem: at each validation point, $\phi_i(v)$ is obtained by inserting $I_i(u), R_i(u, r), C_{-i}(u), S_{-i}(u, r)$ into G. Given the buyer's validation information, provider i can compute $I_i(u)$ and $R_i(u, r)$ exactly from its retained record. The two leave-one-out counts depend on other providers' records and must be estimated under the privacy protocol.

Lemma 3.1. *Write $H_a = \sum_{\ell=1}^{a} 1/\ell$. Fix $I \in \{0, 1\}$ and $R \in \{0, I\}$. We have $|G(I, R, C, S)| \le I/2$. Moreover, for nonnegative integer counts $0 \le S \le C$ and $0 \le S' \le C'$,*

$$\left|G(I, R, C', S') - G(I, R, C, S)\right| \le \frac{2IH_{1+\min\{C,C'\}}}{(1 + \min\{C, C'\})^2}\left(|C' - C| + |S' - S|\right). \tag{7}$$

Lemma 3.1 bounds the change in G by the common factor $2IH_{1+\min\{C,C'\}}/(1 + \min\{C, C'\})^2$ times the linear count-error term $|C' - C| + |S' - S|$, so accurate estimates of both counts give accurate pointwise values.

3.2 HISTOGRAM COUNTS AND PRIVATE ESTIMATION

In TKNN (3), each validation point $v = (u, r)$ defines its own radius-τ neighborhood around u. Privatizing the corresponding indicators separately for every query would accumulate privacy loss

(Kroll, 2021). Instead, a fixed histogram lets each provider send one locally private report that can be reused across queries (Berrett et al., 2021; Györfi & Kroll, 2025; Ma & Yang, 2024; Ma et al., 2026). We construct histogram counterparts of the exact local counts $C_{-i}(u)$ and $S_{-i}(u,r)$, then estimate them from locally private reports. We present the method under the semi-LDP framework of Definition 2.2, which also covers LDP as a special case. For $s > 0$, choose a histogram bandwidth $0 < h \le 1$ with $h^{-1} \in \mathbb{N}$ and partition $[0,1]^s$ into h^{-s} axis-aligned cells A_ℓ^{priv} of side length h, indexed by $\ell \in \{1, \dots, h^{-s}\}$. Assign boundary points deterministically and let g_ℓ be the center of cell ℓ. Let ℓ_k be the index of the cell containing x_k^{priv}. The cell-label histogram has count $\sum_{k=1}^N \mathbf{1}\{\ell_k = \ell, y_k = r\}$ for cell ℓ and label $r \in \{0,1\}$. For $0 < s < d$, these cell indices are combined with each record's exact public coordinates x_k^{pub} when answering a query. Consider a validation point $v = (u, r)$, where $u = (u^{\mathrm{priv}}, u^{\mathrm{pub}}) \in \mathcal{X}$ and $r \in \{0,1\}$. For $a \in [0,1]^{d-s}$, define the histogram neighborhood and its membership indicator by

$$\mathcal{N}_\tau(u,a) = \left\{\ell : \|g_\ell - u^{\mathrm{priv}}\|_2^2 + \|a - u^{\mathrm{pub}}\|_2^2 \le \tau^2\right\}, \qquad I_{k,h}(u) = \mathbf{1}\{\ell_k \in \mathcal{N}_\tau(u, x_k^{\mathrm{pub}})\}.$$

In the private coordinates, ℓ_k locates the histogram cell and its center g_{ℓ_k} replaces x_k^{priv}. This turns neighborhood membership into a selection of fixed histogram cells, with a surrogate private-coordinate distance. Every point in a private cell lies within $\sqrt{s}h/2$ of its center. For Full LDP, we take $h \le \tau$. For $0 < s < d$, we take $h \le \tau/\sqrt{s}$ so this distance error is at most $\tau/2$. Given this neighborhood rule, let $R_{k,h}(u,r) = I_{k,h}(u)\mathbf{1}\{y_k = r\}$ and define the sample counts

$$C_{-i,h}(u) = \sum_{k \ne i} I_{k,h}(u), \qquad S_{-i,h}(u,r) = \sum_{k \ne i} R_{k,h}(u,r), \tag{8}$$

which approximate $C_{-i}(u)$ and $S_{-i}(u,r)$. Each provider sends one private report to estimate the histogram counts in (8). For $s > 0$, provider k releases its public coordinates x_k^{pub} and

$$\widetilde{W}_{k,\ell,r} = \mathbf{1}\{\ell_k = \ell, y_k = r\} + \xi_{k,\ell,r}, \qquad \xi_{k,\ell,r} \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(2/\alpha), \tag{9}$$

for every $\ell \in [h^{-s}]$, $r \in \{0,1\}$. For candidate i and query u, the private versions of (8) are

$$\widetilde{C}_{-i,h}^{\mathrm{raw}}(u) := \sum_{k \ne i} \sum_{\ell \in \mathcal{N}_\tau(u, x_k^{\mathrm{pub}})} \sum_{r'=0}^{1} \widetilde{W}_{k,\ell,r'}, \quad \widetilde{S}_{-i,h}^{\mathrm{raw}}(u,r) := \sum_{k \ne i} \sum_{\ell \in \mathcal{N}_\tau(u, x_k^{\mathrm{pub}})} \widetilde{W}_{k,\ell,r}. \tag{10}$$

We now show that these two statistics satisfy LDP. The final valuation uses only post-processing of these statistics and therefore inherits the same guarantee.

Proposition 3.1 (Privacy of reconstructed counts)**.** *The mechanism in (10) are α-semi-LDP with respect to each provider $k \ne i$. When no features are public ($s = d$), they are α-LDP.*

3.3 Stabilized plug-in valuation

When a neighborhood contains few or no records, privacy noise can dominate the estimated counts in (10) and make the plug-in value unstable (Lemma 3.1). We apply a stabilization trick by mixing the private counts with a positive volume reference. For a real number t, define $\mathrm{trunc}(t,a,b) = \min\{b, \max\{a,t\}\}$. Define the volume-reference count $C_{\mathrm{vol}}(u) := (N-1)h^s \int_{[0,1]^{d-s}} |\mathcal{N}_\tau(u,a)|\, da$, which is $(N-1)$ times the volume of feature locations where a training record would be counted as a neighbor of u under the histogram rule. This quantity can be computed explicitly without privacy concerns. Set the weight $\omega = \min\{1/4, p_-/\sqrt{(N-1)\alpha^2\tau^d h^s}\}$. The cap $1/4$ leaves at least three quarters of the weight

Algorithm 1 PIVOTS (Provider-side private valuation with TKNN-Shapley)

 # User side: private reporting
for each provider k **do**
  Send x_k^{pub} and the statistics in (9).
end for
 # Server side: aggregation and broadcast
Aggregate as in (10). Broadcast these summaries and the validation set $D^{(\mathrm{val})}$.
 # User side: local valuation
for each provider i **do**
  Compute $\widetilde{\Phi}_i$ from z_i, the summaries, and $D^{(\mathrm{val})}$ using (12).
end for

on the truncated private count. The second term shrinks the reference weight as $(N-1)\alpha^2\tau^d h^s$ grows, keeping the squared mixing error at the private count-noise variance scale (Appendix C.1). Here $p_- > 0$ is a conservative public lower bound on the feature density p_X, as specified later in Assumption 4.1. Such a bound may be justified, for example, by prior domain knowledge. We mix the total count and fit the label count as follows:

$$\begin{aligned}\widetilde{C}_{-i,h}(u) &= \Big\lfloor (1-\omega)\,\mathrm{trunc}\big(\widetilde{C}^{\mathrm{raw}}_{-i,h}(u), 0, N-1\big) + \omega C_{\mathrm{vol}}(u) + 1/2 \Big\rfloor, \\ \widetilde{S}_{-i,h}(u,1) &= \operatorname*{arg\,min}_{a\in\{0,\dots,\widetilde{C}_{-i,h}(u)\}} \Big\{ (a - \widetilde{S}^{\mathrm{raw}}_{-i,h}(u,1))^2 + (\widetilde{C}_{-i,h}(u) - a - \widetilde{S}^{\mathrm{raw}}_{-i,h}(u,0))^2 \Big\}. \end{aligned} \tag{11}$$

Set $\widetilde{S}_{-i,h}(u,0) = \widetilde{C}_{-i,h}(u) - \widetilde{S}_{-i,h}(u,1)$. Using these counts, provider i evaluates

$$\widetilde{\phi}_i(v) = G(I_{i,h}(u), R_{i,h}(u,r), \widetilde{C}_{-i,h}(u), \widetilde{S}_{-i,h}(u,r)), \qquad \widetilde{\Phi}_i = M^{-1}\sum_{t=1}^{M} \widetilde{\phi}_i(v_t). \tag{12}$$

The full protocol is summarized in Algorithm 1.

3.4 Privacy of the Buyer Side

The preceding protocol protects providers' records, but the validation information sent to them can reveal the buyer's data. We discuss the alternative choice where the buyer also gets to keep its privacy. We propose to histogramize the validation set using the same partition $\{A^{\mathrm{priv}}_\ell\}$. Depending on the buyer's privacy requirements, this histogram can be released under central DP (Dwork et al., 2006) or LDP (Berrett et al., 2021).

Specifically, let $b(u_t)$ denote the cell containing validation feature u_t, with center $g_{b(u_t)}$. Define its cell-label counts by $V_{b,r} := \sum_{t=1}^{M} \mathbf{1}\{b(u_t) = b\}\mathbf{1}\{r_t = r\}$. If each validation feature is evaluated at its cell center, grouping the pointwise values in (12) by cell and label gives

$$\widetilde{\Phi}_i^{\text{val-hist}} := \frac{1}{M}\sum_{t=1}^{M} \widetilde{\phi}_i\big((g_{b(u_t)}, r_t)\big) = \frac{1}{M}\sum_{b}\sum_{r\in\{0,1\}} V_{b,r}\,\widetilde{\phi}_i\big((g_b, r)\big). \tag{13}$$

Here each $\widetilde{\phi}_i((g_b,r))$ is computed from the provider-side counts by (12). The validation set enters this gridded valuation only through $V_{b,r}$. If $\widetilde{V}_{b,r}$ is a private estimate of $V_{b,r}$, replacing each $V_{b,r}$ by $\widetilde{V}_{b,r}$ in (13) defines $\widetilde{\Phi}_i^{\mathrm{buyer}}$. Since $|G| \le 1/2$ by Lemma 3.1,

$$\big|\widetilde{\Phi}_i^{\mathrm{buyer}} - \widetilde{\Phi}_i^{\text{val-hist}}\big| \le \frac{1}{2M}\sum_{b}\sum_{r\in\{0,1\}} |\widetilde{V}_{b,r} - V_{b,r}|.$$

Thus, an accurate private validation histogram estimate gives an accurate valuation estimate.

4 Theoretical Guarantees

In this section, we establish theoretical guarantees for PIVOTS. We first derive finite-sample mean-squared-error upper bounds. We then give the resulting rates and matching lower bounds. In addition, Appendix C.3 relates valuation error to ranking AUC.

4.1 General upper bound.

For the theoretical analysis, we take $0 < \tau \le 1$ and $0 < \alpha \le \bar{\alpha} < \infty$ for fixed $\bar{\alpha}$, and impose the following sampling and feature-density conditions.

Assumption 4.1 (Sampling and feature densities). The candidate records are i.i.d. from P. The feature marginals of P and Q have densities p_X and q_X on $[0,1]^d$ satisfying $0 < p_- \le p_X(x) \le p_+ < \infty$ and $0 \le q_X(x) \le q_+ < \infty$ a.e. Write $\mathcal{P}$ and $\mathcal{Q}$ for the corresponding distribution classes.

For the realized-pool target $\Phi_i^\star$ in (2), we consider the conditional mean-squared error

$$\mathcal{R}_i := \mathbb{E}\big[(\widetilde{\Phi}_i - \Phi_i^\star)^2 \mid z_i\big].$$

The expectation averages over the other candidate records, private noise, and the validation samples. Let $L_N = \log(e + N\tau^d)$. The following theorem bounds this risk.

Theorem 4.1. *Under Assumption 4.1, for $N \geq 2$ and $M \geq 1$, Algorithm 1 satisfies*

$$\mathcal{R}_i \lesssim \frac{L_N^2}{N^2}\left[\frac{1}{N\alpha^2\tau^d h^s} + \frac{1}{M\tau^d} + \begin{cases} 0, & s = 0, \\ \left(\frac{h}{\tau}\right)^2 + \frac{h}{N\tau^{d+1}}, & s > 0. \end{cases}\right] + \left(\tau^{2d} + \frac{\tau^d}{M}\right) e^{-cN\tau^d}. \quad (14)$$

The upper bound has four terms. The first two terms in (14) apply in all privacy notions. The *private-count error* $1/(N\alpha^2\tau^d h^s)$ captures privacy noise and stabilization. The *validation-sampling error* $1/(M\tau^d)$ comes from averaging over M validation points when only an $O(\tau^d)$ fraction can be neighbors of provider i. When $s > 0$, replacing private features by cell centers shifts neighborhoods by order h, giving the squared *grid bias* $(h/\tau)^2$. Random training records crossing the shifted boundary contribute the *boundary fluctuation* $h/(N\tau^{d+1})$. When $s = 0$, all features are public, so no feature discretization is needed, and neither of these two terms appear. The final remainder $(\tau^{2d} + \tau^d/M)e^{-cN\tau^d}$ arises from unusual events on the sample space and can be exponentially small when τ is not too small.

4.2 Bandwidth selection and resulting rates.

For $s > 0$, write $t = h/\tau$. The leading terms are t^2 and $1/(N\alpha^2\tau^{d+s}t^s)$. Balancing them gives $t \asymp (N\alpha^2\tau^{d+s})^{-1/(s+2)}$. This gives the best bandwidth order within the stated upper bound.

Proposition 4.2 (Rates after bandwidth optimization). *Under the conditions of Theorem 4.1, suppose $M \gtrsim N$ and $N\alpha^2\tau^{d+s} \to \infty$. For $s > 0$, choose $h = \lceil \tau^{-1}(N\alpha^2\tau^{d+s})^{1/(s+2)} \rceil^{-1}$. For all sufficiently large N, Algorithm 1 satisfies*

$$\mathcal{R}_i \lesssim \frac{L_N^2}{N^2}(N\alpha^2\tau^{d+s})^{-2/(s+2)}. \quad (15)$$

The balancing condition implies $h/\tau \to 0$ and $N\tau^d \to \infty$. Thus, the rare-event term becomes negligible automatically. We next establish a matching lower bound, showing that the rate in Proposition 4.2 is optimal for a given τ under the specified reporting protocols.

Theorem 4.3 (Lower bound at the selected resolutions). *Under the conditions of Proposition 4.2, there exists a fixed $z_i \in (0,1)^d \times \{0,1\}$ such that, for all sufficiently large N,*

$$\inf_{\widehat{\Phi}} \sup_{P \in \mathcal{P}, Q \in \mathcal{Q}} \mathbb{E}_{P,Q}[(\widehat{\Phi} - \Phi_i^\star)^2 \mid z_i] \gtrsim \frac{L_N^2}{N^2}(N\alpha^2\tau^{d+s})^{-2/(s+2)}. \quad (16)$$

Here $\widehat{\Phi}$ is any estimator compatible with our privacy setting and recall that $\mathcal{P}$ and $\mathcal{Q}$ are the distribution classes in Assumption 4.1.

The polynomial error decreases as N, α, or τ increases: more records and larger neighborhoods improve count estimation, while a larger privacy budget reduces report noise. For fixed α, τ, d with $M \gtrsim N$, the polynomial N-exponent decreases from 3 at $s = 0$ (Label LDP) through $2 + 2/(s+2)$ for $0 < s < d$ (Semi-LDP) to $2 + 2/(d+2)$ at $s = d$ (LDP), quantifying the cost of privatizing more feature coordinates.

The threshold also matters if the goal is a smooth nonparametric valuation functional rather than the τ-TKNN-Shapley value itself. For $v = (u, r)$, the TKNN utility in (3) averages $\mathbf{1}[Y = r]$ within a radius-τ neighborhood when it is nonempty, just as a local-constant nonparametric regression estimator averages responses to estimate $P(Y = r \mid X = u)$. Under a suitable smoothness assumption, shrinking τ can reduce this regression approximation bias, but fewer neighbors increase the leading private estimation error in (15). To transfer the regression approximation error to a smooth valuation target, one must bound how utility approximation affects the Shapley average over training subsets. With such a bound, choosing τ balances squared approximation error against estimation error, with h chosen conditional on τ. Although the present guarantees address only estimation of the TKNN-Shapley target at a given τ, they provide a foundation for analyzing smooth valuation functionals once the approximation error is controlled.

5 EXPERIMENTS

5.1 SIMULATION STUDY

Simulation setup. We use the 2-dimensional model $X \sim \mathrm{Unif}([0,1]^2)$ and $Y \mid X \sim \mathrm{Bernoulli}(\frac{1}{2} + .05\sum_{j=1}^{2}\sin(2\pi X_j))$. Across label LDP ($s=0$), semi-LDP ($s=1$), and full LDP ($s=2$), we vary $\alpha \in \{1,2,3,4\}$ and $N \in \{2^{11},\ldots,2^{16}\}$ with $M = 1024N$ and $\tau_N = N^{-1/8}$. The validation-size experiment fixes $N = 32768$ and varies M. We report mean squared error (MSE) relative to the realized non-private TKNN-Shapley values, averaged over 200 outer repetitions. Appendix E gives the full experiment details and additional simulation results.

Simulation results. In the first three panels of Figure 2, MSE falls across the entire N range in every disclosure regime. At each displayed (N,α), label LDP has the lowest MSE, semi-LDP lies between label LDP and full LDP, and full LDP has the highest MSE. Increasing α generally shifts the curves downward, although the gain is more modest under full LDP in this setting. The nearly straight log–log N curves show polynomial-like decay and agree qualitatively with the benefit of fewer private feature coordinates in Proposition 4.2. In the rightmost panel, the α curves are close when M is small. Yet, as M increases, they separate and flatten at α-dependent levels. This indicates that validation sampling initially dominates, whereas further validation data have little effect once training-side error dominates, as reflected by the $1/(M\tau^d)$ term in Theorem 4.1.

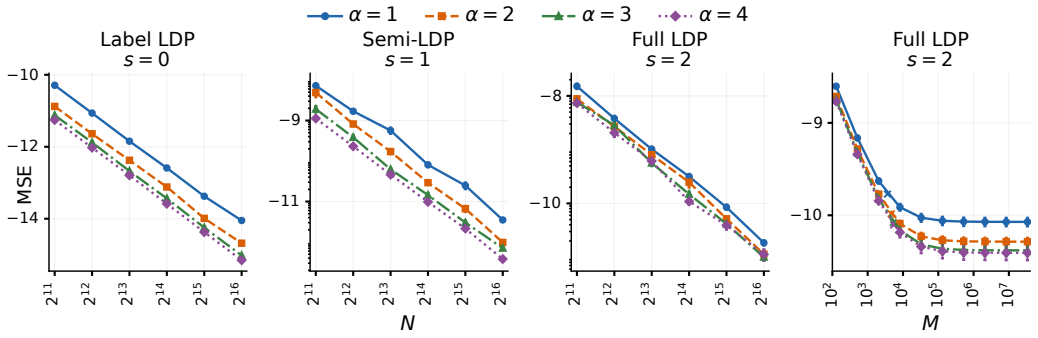

Figure 2: **Simulation results across α, N, M, and s.** Each point shows the mean MSE over 200 repetitions with ± 1 outer standard error bars which are visually negligible.

5.2 REAL-DATA EXPERIMENTS

Setting and tasks. We evaluate private data selection and its effect on downstream prediction on nine binary-classification datasets with 25% label or feature corruption under full LDP, semi-LDP, and label LDP. We compare PIVOTS with a naive method, LDP-KNN, which privatizes records directly before applying KNN-Shapley (Appendix D.1). Both methods use the same data splits and corruption realizations.

For data selection, we measure top-10% overlap between each private ranking and its own non-private counterpart on the same corrupted pool: non-private TKNN-Shapley for PIVOTS and non-private KNN-Shapley for LDP-KNN. We also report corruption-detection AUROC using negative private values as the score, against the known corrupted index. For downstream prediction, we select the top 50% of candidate records by estimated private value and report the resulting classifier's balanced accuracy on clean test data. Table 1 reports $\alpha = 1$. Appendix D gives the experimental protocol and full results for $\alpha \in \{1,2,3,4\}$.

Results. At $\alpha = 1$, PIVOTS has higher mean top-10% recovery and detection AUROC than LDP-KNN on all nine datasets under full LDP for both corruption types (Table 1). Under feature corruption, PIVOTS's top-10% recovery ranges from 14.93% to 26.81%, significantly higher than 0.16% to 9.71% for LDP-KNN.

Table 1: **Real-data results at $\alpha = 1$ (%).** Both corruptions affect 25% of records. KNN denotes LDP-KNN detailed in Appendix D.1. Bold marks the higher private-method mean. Pale blue with $*$ marks a significant difference in a two-sided paired Wilcoxon signed-rank test, after averaging ten privacy runs per split and applying Holm correction across datasets within each comparison (adjusted $p < .05$).

(a) Recovery (own NP-ranking overlap) and detection (AUROC) under full LDP.

	Mislabel (25%)					Feature corruption (25%)				
	Recovery (%)		Detection AUROC (%)			Recovery (%)		Detection AUROC (%)		
Dataset	KNN	PIVOTS	Random	KNN	PIVOTS	KNN	PIVOTS	Random	KNN	PIVOTS
CODRNA	10.69	**21.35***	50.08	62.22	**67.72***	8.93	**22.81***	50.13	51.30	**62.49***
SVMG1	13.32	**21.76***	49.85	85.11	**91.58***	1.96	**20.20***	49.89	59.35	**75.46***
MAGIC	8.70	**19.74***	50.05	61.42	**64.47***	1.61	**20.04***	50.06	42.30	**54.81***
SKIN	14.84	**19.17**	50.47	93.67	**96.95***	4.29	**14.93***	49.88	67.93	**85.03***
PHON	17.28	**26.17***	50.24	72.69	**74.28**	9.71	**20.41***	49.89	47.36	**60.53***
WATER	4.73	**23.30***	50.15	47.75	**51.56***	1.39	**25.58***	50.08	50.12	**50.51**
CALIF	11.05	**23.30***	49.92	62.22	**67.78***	5.51	**26.81***	49.82	50.24	**63.49***
HOUSE8L	5.06	**20.61***	49.85	63.80	**68.24**	0.76	**24.31***	50.05	47.50	**62.64***
PUMP	4.72	**19.65***	50.02	81.73	**83.95**	0.16	**22.45***	49.95	55.72	**77.90***

(b) Downstream balanced accuracy after top-50% selection under full LDP, semi-LDP, and label LDP.

	Mislabel (25%)						Feature corruption (25%)					
	Full LDP		Semi-LDP		Label LDP		Full LDP		Semi-LDP		Label LDP	
Dataset	KNN	PIVOTS	KNN	PIVOTS	KNN	PIVOTS	KNN	PIVOTS	KNN	PIVOTS	KNN	PIVOTS
CODRNA	63.76	**90.40***	64.77	**89.17***	88.99	**91.29**	70.21	**90.58***	74.73	**91.20***	**92.13**	90.52
SVMG1	76.88	**93.62***	77.98	**93.44***	93.59	**93.60**	77.70	**93.60***	77.32	**93.48***	93.10	**93.86**
MAGIC	62.89	**69.23**	67.88	**67.93**	**75.77***	73.97	66.64	**72.35***	70.84	**72.26**	**75.91**	74.31
SKIN	91.99	**94.19**	93.29	**94.77***	92.92	**94.07**	93.29	**94.14**	91.86	**94.23**	91.71	**93.71***
PHON	72.42	**73.90**	71.06	**73.27***	**74.61**	74.56	73.55	**74.27**	72.56	**74.25***	74.39	**75.02**
WATER	50.02	**50.80**	49.96	**50.89**	50.26	**50.47**	50.01	**50.19**	50.31	**50.52**	50.12	**50.53**
CALIF	64.58	**78.26***	66.64	**75.56***	**79.75**	77.70	69.34	**78.88***	70.73	**78.45***	**80.67***	77.91
HOUSE8L	64.41	**74.34***	66.44	**75.17***	**78.72**	77.73	68.65	**77.21***	69.95	**77.72***	**78.75**	77.27
PUMP	77.96	**83.83***	80.20	**80.99**	**86.06**	85.26	80.95	**84.90***	81.88	**83.44**	**86.27***	85.51

After selecting the top 50% of candidate records, PIVOTS yields higher mean accuracy on all nine datasets under both full LDP and semi-LDP for both corruptions, although not every paired difference is significant after Holm correction. Under label LDP, PIVOTS leads on only four of nine downstream datasets under each corruption, and LDP-KNN often has higher mislabel-detection AUROC (Appendix D.3). We attribute the difference to the different valuation rules they use—TKNN-Shapley for PIVOTS and KNN-Shapley for LDP-KNN. Thus, these comparisons do not isolate the effect of private estimation from that of the valuation rule.

6 Conclusion and Discussion

We introduced PIVOTS for provider-side valuation before raw-data disclosure. Each provider sends one locally private report. The buyer aggregates these reports into neighborhood statistics, and providers use the broadcast summary with their retained records to estimate their own values. The framework accommodates varying numbers of private feature coordinates, including classical LDP when all features are private and label LDP when only labels are private.

Our results show that accurate private valuation is possible: under the stated conditions, the estimation error vanishes at a rate bounded above and below for the specified reporting protocols. Simulations exhibit the predicted improvements as the provider sample size or privacy budget increases. On real data, the resulting rankings support corruption detection and value-guided selection, where PIVOTS outperforms the naive LDP-KNN baseline in most of the cases. Our results also quantify the benefit of public features: for fixed privacy budget, neighborhood radius, and ambient dimension, fewer private feature coordinates give a faster asymptotic valuation-error rate.

Our framework also extends privacy protection to the buyer's validation data. If the buyer already holds a validation set, it can release a cell-label histogram under central DP without exposing individual records. Alternatively, validation participants can also submit locally private cell-label reports, so even the buyer does not see their raw data.

AI Assistance Disclosure

In this work, we used generative AI tools to aid and polish writing, produce figures and tables, draft portions of the manuscript, and check mathematical proofs. The authors reviewed all AI-assisted work and take responsibility for the final content, including all text, claims, proofs, and reported results produced with the aid of generative AI.

References

Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. Hadamard response: Estimating distributions privately, efficiently, and with little communication. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1120–1129. PMLR, 2019. URL https://proceedings.mlr.press/v89/acharya19a.html.

Anish Agarwal, Munther Dahleh, and Tuhin Sarkar. A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pp. 701–726. Association for Computing Machinery, 2019. doi: 10.1145/3328526.3329589.

Ashwinkumar Badanidiyuru Varadaraja, Badih Ghazi, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Avinash V. Varadarajan, and Chiyuan Zhang. Optimal unbiased randomizers for regression with label differential privacy. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Raef Bassily. Linear queries estimation with local differential privacy. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 721–729. PMLR, 2019. URL https://proceedings.mlr.press/v89/bassily19a.html.

Thomas B. Berrett, László Györfi, and Harro Walk. Strongly universally consistent nonparametric regression and classification with privatised data. *Electronic Journal of Statistics*, 15(1):2430–2453, 2021.

Robert Istvan Busa-Fekete, Andres Munoz Medina, Umar Syed, and Sergei Vassilvitskii. Label differential privacy and private training data release. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 3233–3251, 2023.

Kamalika Chaudhuri and Daniel Hsu. Sample complexity bounds for differentially private learning. In *Conference on Learning Theory*, volume 19 of *Proceedings of Machine Learning Research*, pp. 155–186, 2011.

Lynn Chua, Qiliang Cui, Badih Ghazi, Charlie Harrison, Pritish Kamath, Walid Krichene, Ravi Kumar, Pasin Manurangsi, Nicolas Mayoraz, Krishna Giri Narra, Steffen Rendle, Amer Sinha, Avinash Varadarajan, and Chiyuan Zhang. Training differentially private ad prediction models with semi-sensitive features. In *AdKDD Workshop at KDD*, 2024.

Mihaela Curmei, Walid Krichene, Li Zhang, and Mukund Sundararajan. Private matrix factorization with public item features. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pp. 805–812, 2023.

John C. Duchi, Michael I. Jordan, and Martin J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018. doi: 10.1080/01621459.2017.1389735.

Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, volume 3876 of *Lecture Notes in Computer Science*, pp. 265–284. Springer, 2006. doi: 10.1007/11681878_14.

Badih Ghazi, Noah Golowich, Ravi Kumar, Pasin Manurangsi, and Chiyuan Zhang. Deep learning with label differential privacy. In *Advances in Neural Information Processing Systems*, volume 34, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/e3a54649aeec04cf1c13907bc6c5c8aa-Abstract.html.

Badih Ghazi, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Avinash V. Varadarajan, and Chiyuan Zhang. Regression with label differential privacy. In *International Conference on Learning Representations*, 2023.

Badih Ghazi, Yangsibo Huang, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, and Chiyuan Zhang. LabelDP-Pro: Learning with label differential privacy via projections. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/00153f90d9177dd3a872133972bc8dea-Abstract-Conference.html.

Amirata Ghorbani and James Zou. Data Shapley: Equitable valuation of data for machine learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2242–2251. PMLR, 2019. URL https://proceedings.mlr.press/v97/ghorbani19c.html.

Amirata Ghorbani, Michael P. Kim, and James Zou. A distributional framework for data valuation. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 3535–3544. PMLR, 2020. URL https://proceedings.mlr.press/v119/ghorbani20a.html.

László Györfi and Martin Kroll. On rate optimal private regression under local differential privacy. *Statistica Sinica*, 35:613–627, 2025. doi: 10.5705/ss.202022.0186.

Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nezihe Merve Gürel, Bo Li, Ce Zhang, Costas J. Spanos, and Dawn Song. Efficient task-specific data valuation for nearest neighbor algorithms. *Proceedings of the VLDB Endowment*, 12(11):1610–1623, 2019a. doi: 10.14778/3342263.3342637.

Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, Dawn Song, and Costas J. Spanos. Towards efficient data valuation based on the Shapley value. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1167–1176. PMLR, 2019b. URL https://proceedings.mlr.press/v89/jia19a.html.

Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2436–2444. PMLR, 2016a. URL https://proceedings.mlr.press/v48/kairouz16.html.

Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *Journal of Machine Learning Research*, 17(17):1–51, 2016b.

Walid Krichene, Nicolas Mayoraz, Steffen Rendle, Shuang Song, Abhradeep Thakurta, and Li Zhang. Private learning with public features. In *International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 4150–4158, 2024.

Martin Kroll. On density estimation at a fixed point under local differential privacy. *Electronic Journal of Statistics*, 15(1):1783–1813, 2021. doi: 10.1214/21-EJS1830.

Yongchan Kwon and James Zou. Beta Shapley: A unified and noise-reduced data valuation framework for machine learning. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 8780–8802. PMLR, 2022. URL https://proceedings.mlr.press/v151/kwon22a.html.

Yongchan Kwon and James Zou. Data-OOB: Out-of-bag estimate as a simple and efficient data value. In *International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 18135–18152, 2023.

Yuheng Ma and Hanfang Yang. Optimal locally private nonparametric classification with public data. *Journal of Machine Learning Research*, 25(167):1–62, 2024.

Yuheng Ma, Hanfang Yang, and Ke Jia. Locally private estimation with public features. *Journal of Machine Learning Research*, 27(183):1–58, 2026. URL https://www.jmlr.org/papers/v27/25-1896.html.

Mani Malek Esmaeili, Ilya Mironov, Karthik Prasad, Igor Shilov, and Florian Tramèr. Antipodes of label differential privacy: PATE and ALIBI. In *Advances in Neural Information Processing Systems*, volume 34, pp. 6934–6945, 2021.

Li Peng, Jiayao Zhang, Yihang Wu, Weiran Liu, Jinfei Liu, Zheng Yan, Kui Ren, Lei Zhang, and Lin Qu. Reliable and private utility signaling for data markets. *Proceedings of the ACM on Management of Data*, 3(6):1–27, 2025. doi: 10.1145/3769821.

Zeyu Shen, Anilesh Krishnaswamy, Janardhan Kulkarni, and Kamesh Munagala. Classification with partially private features. *arXiv preprint arXiv:2312.07583*, 2023.

Rachael Sim, Yehong Zhang, Nghia Hoang, Xinyi Xu, Bryan Kian Hsiang Low, and Patrick Jaillet. Incentives in private collaborative machine learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 7555–7593, 2023. doi: 10.52202/075280-0333.

Zhihua Tian, Jian Liu, Jingyu Li, Xinle Cao, Ruoxi Jia, Jun Kong, Mengdi Liu, and Kui Ren. Private data valuation and fair payment in data marketplaces. *arXiv preprint arXiv:2210.08723*, 2022. URL https://arxiv.org/abs/2210.08723.

Di Wang and Jinhui Xu. On sparse linear regression in the local differential privacy model. In *International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 6628–6637, 2019.

Jiachen T. Wang and Ruoxi Jia. Data Banzhaf: A robust data valuation framework for machine learning. In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pp. 6388–6421. PMLR, 2023a. URL https://proceedings.mlr.press/v206/wang23e.html.

Jiachen T. Wang and Ruoxi Jia. A note on "efficient task-specific data valuation for nearest neighbor algorithms". *arXiv preprint arXiv:2304.04258*, 2023b.

Jiachen (Tianhao) Wang, Yuqing Zhu, Yu-Xiang Wang, Ruoxi Jia, and Prateek Mittal. A privacy-friendly approach to data valuation. In *Advances in Neural Information Processing Systems*, volume 36, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/bdd5522a32b3a959a6d81fb6ddc1cb38-Abstract-Conference.html.

Jiachen (Tianhao) Wang, Prateek Mittal, Dawn Song, and Ruoxi Jia. Data Shapley in one training run. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/20fdaf67581e6d7157376d1ed584040a-Abstract-Conference.html.

Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. Locally differentially private protocols for frequency estimation. In *26th USENIX Security Symposium (USENIX Security 17)*, pp. 729–745. USENIX Association, 2017. URL https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/wang-tianhao.

Mimee Xu, Laurens van der Maaten, and Awni Hannun. Data appraisal without data sharing. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pp. 11422–11437. PMLR, 2022.

Shirong Xu, Chendi Wang, Will Wei Sun, and Guang Cheng. Binary classification under local label differential privacy using randomized response mechanisms. *Transactions on Machine Learning Research*, 2023.

Puning Zhao, Chuan Ma, Li Shen, Shaowei Wang, and Rongfei Fan. On theoretical limits of learning with label differential privacy. *arXiv preprint arXiv:2502.14309*, 2025a.

Puning Zhao, Jiafei Wu, Zhe Liu, Li Shen, Zhikun Zhang, Rongfei Fan, Le Sun, and Qingming Li. Enhancing learning with label differential privacy by vector approximation. In *International Conference on Learning Representations*, 2025b.

Shuyuan Zheng, Yang Cao, and Masatoshi Yoshikawa. Secure Shapley value for cross-silo federated learning. *Proceedings of the VLDB Endowment*, 16(7):1657–1670, 2023. doi: 10.14778/3587136.3587141.

APPENDIX CONTENTS

A Protocol Details and Privacy Guarantees **15**

A.1 Full LDP: the original full-dimensional Laplace report 16

A.2 Semi-LDP: private bins and exact public coordinates 16

A.3 Label LDP: private label channels and exact neighborhoods 17

A.4 Feasible fitting, stabilization, and local evaluation 17

B Local-count representation and stability **18**

C Detailed Statistical Guarantees and Proofs **19**

C.1 Proof of the upper bounds 20

C.2 Proof of the matching lower bounds 25

C.3 From valuation error to ranking AUC 28

D Details for Real-Data Experiments **29**

D.1 Protocol and Reporting 29

D.2 Recovery 31

D.3 Detection 34

D.4 Downstream prediction 37

E Additional Simulation Results **43**

E.1 Shared setup 43

E.2 Risk and uncertainty 43

E.3 Validation-size plateau 43

E.4 One-dimensional results 43

E.5 Controlled count ablation 44

E.6 Scaling without stabilization 46

F Extended Related Work **47**

F.1 Data Valuation and Tractable Shapley Computation 47

F.2 Privacy-Preserving Valuation and Pre-Acquisition Appraisal 47

F.3 Locally Private Statistical Estimation 47

F.4 Public Features, Partial Privacy, and Label Privacy 48

A Protocol Details and Privacy Guarantees

Let $D = \{z_k\}_{k=1}^N$, where $z_k = (x_k, y_k) \in [0,1]^d \times \{0,1\}$. Fix a provider i and its retained record z_i. Write $x = (x^{\mathrm{priv}}, x^{\mathrm{pub}})$ for a fixed, publicly known coordinate split, with s private coordinates and $d - s$ public coordinates. Here $s = d$ gives Full LDP, $0 < s < d$ gives Semi-LDP, and $s = 0$ gives Label LDP. In particular, the semi-LDP construction below applies to every integer $d \geq 2$ and $1 \leq s \leq d-1$. The label is protected in every regime. The designated-public-feature interpretation agrees with the model of Ma et al. (2026).

All preprocessing, coordinate splits, thresholds, grids, and tie rules are deterministic public design choices fixed before reporting. All distances below are computed in the same publicly specified coordinate system. The validation sample $D^{(\mathrm{val})} = \{v_t\}_{t=1}^M$ is non-sensitive. Write a validation record as $v = (u, r)$, where $u \in [0,1]^d$ and $r \in \{0,1\}$.

For a fixed threshold $\tau > 0$, let

$$\begin{aligned} I_k(u) &= \mathbf{1}\{\|x_k - u\|_2 \leq \tau\}, \qquad R_k(u,r) = I_k(u)\mathbf{1}\{y_k = r\}, \\ C_{-i}(u) &= \sum_{k \neq i} I_k(u), \qquad\qquad S_{-i}(u,r) = \sum_{k \neq i} R_k(u,r). \end{aligned} \tag{17}$$

The subscript $-i$ indicates that provider i is excluded. A bar denotes the expectation over D_{-i} conditional on z_i; in particular, $\overline{C}_{-i}(u) = (N-1)P_X(\|X - u\|_2 \leq \tau)$ and $\overline{S}_{-i}(u,r) = (N-1)P(\|X-u\|_2 \leq \tau, Y = r)$. For the algebraic bounds below, write the sum in (6) equivalently using $H_a = \sum_{j=1}^a j^{-1}$:

$$G(I,R,C,S) = \begin{cases} R - I/2, & C = 0, \\ \dfrac{RH_{C+1} - I\{(S/C)(H_{C+1} - 1) + 1/2\}}{C+1}, & C \geq 1. \end{cases} \tag{18}$$

The exact TKNN-Shapley value at v is $\phi_i(v) = G(I_i(u), R_i(u,r), C_{-i}(u), S_{-i}(u,r))$, using the empty-coalition utility $1/2$. This is a local-count form of the TKNN-Shapley rule (Wang et al., 2023). The target remains

$$\Phi_i^\star = \int \phi_i(v)\, Q(dv). \tag{19}$$

For each realized D, this integral averages over the validation population only. The target is not replaced by its expectation over D.

Assumption A.1 (Sampling and feature densities)**.** Conditional on the fixed z_i, the other $N-1$ records are i.i.d. from P. The M validation records are i.i.d. from Q, independently of the training records and reporting randomness. The feature densities satisfy

$$0 < p_- \leq p_X \leq p_+ < \infty, \qquad 0 \leq q_X \leq q_+ < \infty \quad \text{a.e. on } [0,1]^d.$$

We take $p_- \leq 1 \leq p_+$ and $q_+ \geq 1$, as required for the corresponding density classes to be nonempty. No smoothness of either conditional label distribution is assumed. As in Assumption 4.1, denote these classes by $\mathcal{P}$ and $\mathcal{Q}$.

The lower bound on p_X makes the expected exact and approximate neighborhood counts comparable to $(N-1)\tau^d$, including at the cube boundary. No regularity of $P(Y = 1 \mid X = x)$ is needed: the geometric approximation moves a neighborhood boundary rather than smoothing the conditional label function.

Throughout, d, p_-, p_+, q_+ and $\bar{\alpha} < \infty$ are fixed, $0 < \alpha \leq \bar{\alpha}$, $0 < \tau \leq 1$, $N \geq 2$, and $M \geq 1$. All implicit constants depend only on these fixed quantities. For Full LDP assume $h^{-1} \in \mathbb{N}$ and $0 < h \leq \tau$. For the exact-query Semi-LDP construction assume

$$h^{-1} \in \mathbb{N}, \qquad 0 < h \leq \frac{\tau}{\sqrt{s}}. \tag{20}$$

This last condition is a sufficient geometric resolution condition, not a privacy requirement or a restriction on d and s. It guarantees uniformly positive approximate-neighborhood volume even at the boundary. It holds eventually for the rate-selected grid below and therefore does not change the asymptotic exponent. Label LDP uses no grid.

A.1 Full LDP: the original full-dimensional Laplace report

Partition $[0,1]^d$ into regular cells A_j of side length h, with centers g_j and a deterministic boundary assignment. Let j_k be the cell of x_k, $b = b(u)$ the cell of u, and

$$\mathcal{N}_\tau(b) = \{j : \|g_j - g_b\|_2 \le \tau\}.$$

Provider k sends

$$\widetilde{W}_{k,j,r} = \mathbf{1}\{j_k = j, y_k = r\} + \xi_{k,j,r}, \qquad \xi_{k,j,r} \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(2/\alpha), \qquad r \in \{0,1\}. \tag{21}$$

Every coordinate is randomized, including zero-signal coordinates. The buyer forms $\widetilde{T}_{j,r} = \sum_k \widetilde{W}_{k,j,r}$, and provider i computes

$$\widetilde{Y}_r(u) = \sum_{j \in \mathcal{N}_\tau(b)} (\widetilde{T}_{j,r} - \widetilde{W}_{i,j,r}), \qquad I_{k,h}(u) = \mathbf{1}\{j_k \in \mathcal{N}_\tau(b)\}. \tag{22}$$

Thus the Full-LDP reporting and neighborhood conventions are unchanged.

A.2 Semi-LDP: private bins and exact public coordinates

Partition $[0,1]^s$ into regular private cells A_ℓ^{priv} of side h and centers g_ℓ. Let ℓ_k index the private cell containing x_k^{priv}. Provider k sends its exact x_k^{pub} and the vector

$$\widetilde{W}_{k,\ell,r} = \mathbf{1}\{\ell_k = \ell, y_k = r\} + \xi_{k,\ell,r}, \qquad \xi_{k,\ell,r} \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(2/\alpha). \tag{23}$$

No public coordinate is binned. The validation coordinate u is also used exactly. For an exact public vector $a \in [0,1]^{d-s}$, define

$$\mathcal{N}_\tau(u,a) = \left\{\ell : \|g_\ell - u^{\mathrm{priv}}\|_2^2 + \|a - u^{\mathrm{pub}}\|_2^2 \le \tau^2\right\}. \tag{24}$$

The buyer evaluates query-specific aggregates

$$\widetilde{T}_r(u) = \sum_{k=1}^{N} \sum_{\ell \in \mathcal{N}_\tau(u, x_k^{\mathrm{pub}})} \widetilde{W}_{k,\ell,r}. \tag{25}$$

Provider i removes its entire corresponding report contribution:

$$\widetilde{Y}_r(u) = \widetilde{T}_r(u) - \sum_{\ell \in \mathcal{N}_\tau(u, x_i^{\mathrm{pub}})} \widetilde{W}_{i,\ell,r}, \tag{26}$$

$$I_{k,h}(u) = \mathbf{1}\Big\{\|g_{\ell_k} - u^{\mathrm{priv}}\|_2^2 + \|x_k^{\mathrm{pub}} - u^{\mathrm{pub}}\|_2^2 \le \tau^2\Big\}. \tag{27}$$

The latent index ℓ_k is used only to define the unobserved noise-free count and the provider's own indicator. The buyer does not observe ℓ_k: it computes (25) by summing all selected coordinates of the noisy vector.

For both feature-private regimes, write

$$R_{k,h}(u,r) = I_{k,h}(u)\mathbf{1}\{y_k = r\}, \qquad C_{-i,h}(u) = \sum_{k \ne i} I_{k,h}(u), \qquad S_{-i,h}(u,r) = \sum_{k \ne i} R_{k,h}(u,r). \tag{28}$$

Complete self-report subtraction gives

$$\mathbb{E}_W[\widetilde{Y}_r(u) \mid D] = S_{-i,h}(u,r). \tag{29}$$

Removing only the true self-label, rather than the complete selected noisy self-report, would not implement this estimator.

A.3 LABEL LDP: PRIVATE LABEL CHANNELS AND EXACT NEIGHBORHOODS

Set $p_\alpha = (1 + e^\alpha)^{-1}$. Independently across providers, send $(x_k, \widetilde{y}_k)$, where

$$\Pr(\widetilde{y}_k = y_k \mid z_k) = 1 - p_\alpha, \qquad \Pr(\widetilde{y}_k = 1 - y_k \mid z_k) = p_\alpha. \tag{30}$$

This is binary randomized response; the frequency-reconstruction principle is standard in LDP estimation (Wang et al., 2017). Define the post-processed channels

$$\widetilde{W}_{k,1} = \frac{\widetilde{y}_k - p_\alpha}{1 - 2p_\alpha}, \qquad \widetilde{W}_{k,0} = 1 - \widetilde{W}_{k,1}. \tag{31}$$

The alternative Laplace implementation releases

$$\widetilde{W}_{k,1} = y_k + \xi_k, \qquad \xi_k \overset{\text{i.i.d.}}{\sim} \mathrm{Lap}(1/\alpha), \qquad \widetilde{W}_{k,0} = 1 - \widetilde{W}_{k,1}. \tag{32}$$

Choose one of these mechanisms before reporting, and use its channels in the same reconstruction and fit below. Do not clip the individual channels before aggregation. For bookkeeping only, put $I_{k,h} = I_k$, $R_{k,h} = R_k$, $C_{-i,h} = C_{-i}$, and $S_{-i,h} = S_{-i}$ in this regime; no bandwidth is used. Reconstruct

$$\widetilde{Y}_r(u) = \sum_{k \neq i} I_k(u) \widetilde{W}_{k,r}, \qquad \widetilde{Y}_0(u) + \widetilde{Y}_1(u) = C_{-i}(u). \tag{33}$$

In particular,

$$\begin{aligned} \mathbb{E}_W[\widetilde{Y}_1(u) \mid D] &= S_{-i}(u,1), \\ \mathrm{Var}_W(\widetilde{Y}_1(u) \mid D) &= C_{-i}(u) \begin{cases} e^\alpha/(e^\alpha - 1)^2, & \text{randomized response,} \\ 2/\alpha^2, & \text{Laplace report.} \end{cases} \end{aligned} \tag{34}$$

The indicator and the total count are exact; only the label count is privately reconstructed.

A.4 FEASIBLE FITTING, STABILIZATION, AND LOCAL EVALUATION

Let $B_h(u) = \{x : I_h(x,u) = 1\}$ denote the approximate feature neighborhood specified above. For $s > 0$, retain the same geometric-reference idea, but use the volume of the actual neighborhood:

$$C_{\mathrm{vol}}(u) = (N-1)\,\mathrm{Leb}_d(B_h(u)), \qquad \omega = \min\left\{\frac{1}{4}, \frac{p_-}{\sqrt{(N-1)\alpha^2\tau^d h^s}}\right\}. \tag{35}$$

For Full LDP this is exactly $C_{\mathrm{vol}}(u) = (N-1)|\mathcal{N}_\tau(b)|h^d$. For Semi-LDP it is instead

$$\begin{aligned} C_{\mathrm{vol}}(u) &= (N-1)h^s \int_{[0,1]^{d-s}} |\mathcal{N}_\tau(u,a)|\,da \\ &= (N-1)h^s \sum_\ell \mathrm{Leb}_{d-s}\left\{a \in [0,1]^{d-s} : \|g_\ell - u^{\mathrm{priv}}\|_2^2 + \|a - u^{\mathrm{pub}}\|_2^2 \le \tau^2\right\}. \end{aligned} \tag{36}$$

This reference is deterministic at a fixed query and equals the expected neighborhood count only under a uniform feature density. It differs in general from the population means $\overline{C}_{-i,h}(u) = \mathbb{E}_{D_{-i}} C_{-i,h}(u) = (N-1)P_X(B_h(u))$ and $\overline{S}_{-i,h}(u,r) = \mathbb{E}_{D_{-i}} S_{-i,h}(u,r) = (N-1)P(X \in B_h(u), Y = r)$. The theorem concerns this exact geometric reference, not an unspecified numerical approximation to it. The reference and the expected true count are both of order $\mu = (N-1)\tau^d$, although they need not be equal under a nonuniform feature density. Indeed, $p_- C_{\mathrm{vol}}(u) \le \overline{C}_{-i,h}(u) \le p_+ C_{\mathrm{vol}}(u)$. Mixing therefore introduces bias. The factor p_- makes the weight conservative for the specified density class; with the chosen ω, its contribution to mean squared count error is at most of order $\mu/(\alpha^2 h^s)$, matching the raw report-noise order in (58); see (61). The same geometric reference limits the effect of rare severe underestimation of the truncated raw total.

The raw channel $\widetilde{Y}_r(u)$ here equals $\widetilde{S}^{\mathrm{raw}}_{-i,h}(u,r)$ in the main text. Suppress u below. For $s > 0$, truncate the raw total before mixing:

$$\widetilde{C}_{-i,h} = \left\lfloor (1-\omega)\,\mathrm{trunc}(\widetilde{Y}_0 + \widetilde{Y}_1, 0, N-1) + \omega C_{\mathrm{vol}} + \frac{1}{2} \right\rfloor, \tag{37}$$

$$\widetilde{S}_{-i,h}(1) \in \operatorname*{arg\,min}_{a \in \{0,\dots,\widetilde{C}_{-i,h}\}} \left[(a - \widetilde{Y}_1)^2 + (\widetilde{C}_{-i,h} - a - \widetilde{Y}_0)^2\right], \qquad \widetilde{S}_{-i,h}(0) = \widetilde{C}_{-i,h} - \widetilde{S}_{-i,h}(1). \tag{38}$$

Ties in (38) use the smaller a. Since both the truncated raw total and C_{vol} lie in $[0, N-1]$, all final counts are feasible. For Label LDP, set $\widetilde{C}_{-i,h} = C_{-i}$ and apply only (38). Its objective is twice $(a - \widetilde{Y}_1)^2$, since the two raw channels sum to C_{-i}. There is no total-count mixing in this regime.

Each provider locally evaluates

$$\widetilde{\phi}_i(u,r) = G\big(I_{i,h}(u), R_{i,h}(u,r), \widetilde{C}_{-i,h}(u), \widetilde{S}_{-i,h}(u,r)\big), \qquad \widetilde{\Phi}_i = \frac{1}{M}\sum_{t=1}^{M} \widetilde{\phi}_i(v_t). \tag{39}$$

The $C = 0$ branch of G is always used when needed; no query is discarded.

Broadcast and information boundary. Under Full LDP the original training aggregate and validation histogram suffice. Under Semi-LDP the buyer instead broadcasts each non-sensitive validation record (u, r) and its query aggregates $(\widetilde{T}_0(u), \widetilde{T}_1(u))$; each provider performs (26) locally. Under Label LDP the buyer broadcasts (u, r), $\sum_k I_k(u)$, and $\sum_k I_k(u)\widetilde{W}_{k,1}$. Provider i subtracts $I_i(u)$ and $I_i(u)\widetilde{W}_{i,1}$, respectively. A coarse validation histogram alone does not implement exact-query Semi-LDP or Label evaluation. The same once-generated reports are reused for all queries. The modified query broadcasts do not claim the original grid-histogram communication cost or a constant communication cost.

Proof of Proposition 3.1. For $s > 0$, the unnoised cell–label report is one-hot. Two such vectors with the same public features have ℓ_1 distance at most 2, so adding independent $\mathrm{Lap}(2/\alpha)$ noise to every coordinate gives a report-density ratio at most e^{α}. This includes the Full-LDP endpoint $s = d$, where there are no public features. For fixed i and $k \neq i$, the entire collection of reconstructed counts is a post-processing of k's report and the other providers' independent reports, using only public design and non-sensitive validation queries. It therefore inherits the same bound, without additional privacy loss across queries. For the Label-LDP endpoint $s = 0$, binary randomized response has likelihood ratio at most e^{α}; the alternative Laplace label channel has sensitivity 1 and the same bound. Conditional on z_i, the local valuation $\widetilde{\Phi}_i$ also inherits the bound with respect to each $k \neq i$ by post-processing. □

The valuation uses z_i without randomization and stays local to provider i; releasing it does not, by this proposition alone, protect i's own record.

B LOCAL-COUNT REPRESENTATION AND STABILITY

Lemma B.1 (Representation and count stability)**.** *The exact TKNN-Shapley value is given by (18). For $I \in \{0,1\}$, $R \in \{0, I\}$ and feasible nonnegative integer counts,*

$$|G(I,R,C,S)| \le I\frac{H_{C+1} - 1/2}{C+1} \le \frac{I}{2}, \tag{40}$$

$$|G(I,R,C',S') - G(I,R,C,S)| \le \frac{2IH_{1+\min\{C,C'\}}}{(1+\min\{C,C'\})^2}(|C'-C| + |S'-S|). \tag{41}$$

Proof. A non-neighbor is a dummy player. Restricting a uniformly random player permutation to provider i and its $C_{-i}(u)$ other threshold neighbors gives a uniform permutation of these $C_{-i}(u)+1$ players. If $I_i(u) = 0$ the value is zero. If $I_i(u) = 1$, provider i is equally likely to occupy any of the $C_{-i}(u)+1$ positions. With no preceding neighbor, the marginal contribution is $R_i(u,r) - 1/2$. With $m \ge 1$ preceding neighbors, their expected fraction of matching labels is $S_{-i}(u,r)/C_{-i}(u)$, so the expected marginal is $(R_i(u,r) - S_{-i}(u,r)/C_{-i}(u))/(m+1)$. Consequently, when $C_{-i}(u) \ge 1$,

$$\phi_i(v) = \frac{1}{C_{-i}(u)+1}\left[R_i(u,r) - \frac{1}{2} + \left(R_i(u,r) - \frac{S_{-i}(u,r)}{C_{-i}(u)}\right)\sum_{m=1}^{C_{-i}(u)}\frac{1}{m+1}\right],$$

which is (18); $C_{-i}(u) = 0$ gives the other branch. In the remaining algebra, C, S denote generic inputs of G.

The case $I = 0$ is immediate. For $I = 1$, maximize the absolute value of (18) over $R \in \{0, 1\}$ and $S \in [0, C]$ to obtain the first inequality in (40). The second follows from $H_{C+1} - 1 \le C/2$. For $C \ge 1$,

$$G(I, R, C, S+1) - G(I, R, C, S) = -I\frac{H_{C+1} - 1}{C(C+1)},$$
$$G(I, R, C+1, S) - G(I, R, C, S) = -\frac{G(I, R, C, S)}{C+2} + \frac{R - IS/(C+1)}{(C+2)^2} + \frac{IS(H_{C+1} - 1)}{C(C+1)(C+2)}.$$

The first increment has magnitude at most $IH_{C+1}/(C+1)^2$ because $H_{C+1} \le C+1$. Using (40), $S \le C$, and $|R - IS/(C+1)| \le I$ bounds the second increment by

$$\frac{I(2H_{C+1} - 1)}{(C+1)(C+2)} + \frac{I}{(C+2)^2} \le \frac{2IH_{C+1}}{(C+1)^2}.$$

At $C = 0$, necessarily $S = 0$, and the increment to $(1, 0)$ is $(I - R)/4$, which also satisfies this bound. If $C \le C'$, first increase C to C' while holding S fixed, then move S to S' at total C'. Every intermediate state is feasible. The sequence $H_{a+1}/(a+1)^2$ is decreasing, so telescoping proves (41). Exchange the two pairs when $C > C'$. □

C Detailed Statistical Guarantees and Proofs

Write $L_N = \log(e + N\tau^d)$ and $\mathcal{R}_i = \mathbb{E}[(\widetilde{\Phi}_i - \Phi_i^\star)^2 \mid z_i]$. The expectation includes the other training records, reports, and validation sample. It does not average the training dataset out of the target.

Theorem C.1 (Upper bounds for the specified reports)**.** *Under Assumption A.1, the Full LDP and Semi-LDP estimators, with $s > 0$, satisfy*

$$\mathcal{R}_i \lesssim \frac{L_N^2}{N^2}\left[\left(\frac{h}{\tau}\right)^2 + \frac{h}{N\tau^{d+1}} + \frac{1}{N\alpha^2\tau^d h^s} + \frac{1}{M\tau^d}\right] + \left(\tau^{2d} + \frac{\tau^d}{M}\right)e^{-cN\tau^d}. \tag{42}$$

The exact-neighborhood Label estimator, using either randomized response or the Laplace label report, satisfies

$$\mathcal{R}_i \lesssim \frac{L_N^2}{N^2}\left[\frac{1}{N\alpha^2\tau^d} + \frac{1}{M\tau^d}\right] + \left(\tau^{2d} + \frac{\tau^d}{M}\right)e^{-cN\tau^d}. \tag{43}$$

No discretization term appears for Label LDP. The constant $c > 0$ has only the allowed dependence.

Corollary C.2 (Rates at the selected resolutions)**.** *Under the conditions of Theorem C.1, suppose $M \gtrsim N$ and $N\alpha^2\tau^{d+s} \to \infty$. All conclusions hold for sufficiently large N. For $s > 0$, choose*

$$h_0 = \tau(N\alpha^2\tau^{d+s})^{-1/(s+2)}, \qquad h = 1/\lceil h_0^{-1} \rceil.$$

Label LDP has no grid. Then

$$\mathcal{R}_i \lesssim \frac{L_N^2}{N^2}(N\alpha^2\tau^{d+s})^{-2/(s+2)}. \tag{44}$$

In particular, Full LDP, Semi-LDP, and Label LDP correspond respectively to

$$\frac{L_N^2}{N^2}(N\alpha^2\tau^{2d})^{-2/(d+2)}, \qquad \frac{L_N^2}{N^2}(N\alpha^2\tau^{d+s})^{-2/(s+2)}, \qquad \frac{L_N^2}{N^2}(N\alpha^2\tau^d)^{-1}.$$

Theorem C.3 (Lower bounds for the prescribed reporting experiments)**.** *Fix one of the reports above and, for $s > 0$, its admissible private grid. There is a fixed $z_i \in (0, 1)^d \times \{0, 1\}$ such that, whenever $(N-1)\tau^d$ exceeds a sufficiently large fixed constant, for every $M \ge 1$,*

$$\inf_T \sup_{P \in \mathcal{P},\, Q \in \mathcal{Q}} \mathbb{E}_{P,Q}\left[(T - \Phi_i^\star)^2 \mid z_i\right] \gtrsim \frac{L_N^2}{N^2}\min\left\{1, \frac{1}{N\alpha^2\tau^d h^s}\right\}. \tag{45}$$

Here $\mathcal{P}, \mathcal{Q}$ are the classes in Assumption A.1; when $s = 0$, the notation h^s means 1 and does not introduce a grid. The estimator may use all individual prescribed reports, designated public features, its observed z_i, and the validation sample, rather than only the buyer's aggregates.

Under the growth and resolution conditions of Corollary C.2, the right-hand side of (45) is of order

$$\frac{L_N^2}{N^2}(N\alpha^2\tau^{d+s})^{-2/(s+2)}. \tag{46}$$

Thus the leading rates match for these specified reporting protocols at the selected resolutions when $M \gtrsim N$. This is not a lower bound for each finite-sample upper-bound term, including validation sampling, and does not optimize over other reporting mechanisms or, in the feature-private cases, other grids.

C.1 Proof of the upper bounds

Proof of Theorem C.1. Fix z_i and suppress it from expectations. Let

$$\mu = (N-1)\tau^d, \qquad c_0 = p_-\, 2^{-d} d^{-d/2}.$$

The extra 2^{-d} relative to the full-grid volume bound accommodates the exact-query semi-LDP geometry. Constants denoted by $K, c > 0$ may change between lines and have only the dependence stated above. When the query u is fixed, we omit it from count arguments where convenient. Let $\widehat{\phi}_{i,h}(v) = G(I_{i,h}, R_{i,h}, C_{-i,h}, S_{-i,h}(r))$. Under Label LDP, $\widehat{\phi}_{i,h} = \phi_i$ identically.

Step 1: Geometry, support, and boundary counts. Write $B(u,t) = \{x \in [0,1]^d : \|x-u\|_2 \le t\}$ and let v_d be the volume of the Euclidean unit ball. The box with coordinate half-width $\tau/\sqrt{d}$ centered at u, intersected with the cube, is contained in $B(u,\tau)$. Every side has length at least $\tau/\sqrt{d}$, so

$$d^{-d/2}\tau^d \le \mathrm{Leb}_d(B(u,\tau)) \le v_d\tau^d. \tag{47}$$

For Full LDP put $q = h^{-1}$ and identify cells with indices in $\{0,\dots,q-1\}^d$. At an index b_j, the number of indices within integer distance t in one coordinate is

$$1 + \min\{t, b_j\} + \min\{t, q-1-b_j\} \ge \min\{q, t+1\}.$$

It is also at most $2t+1$. Hence, with $J_b = |\mathcal{N}_\tau(b)|$,

$$\left[\min\left\{q, \left\lfloor \frac{\tau}{\sqrt{d}\,h}\right\rfloor + 1\right\}\right]^d \le J_b \le (2\lfloor \tau/h\rfloor + 1)^d.$$

Both entries of the minimum are at least $\tau/(\sqrt{d}\,h)$, and $2\lfloor\tau/h\rfloor + 1 \le 3\tau/h$. Therefore

$$d^{-d/2}\tau^d \le J_b h^d = \mathrm{Leb}_d(B_h(u)) \le 3^d\tau^d. \tag{48}$$

This argument includes boundary cells.

For Semi-LDP, replacing only x^{priv} by its cell center moves the feature vector by at most $\sqrt{s}h/2$. Consequently,

$$\left|\left\|(g_{\ell(x)}, x^{\mathrm{pub}}) - u\right\|_2 - \|x-u\|_2\right| \le \frac{\sqrt{s}h}{2}, \qquad B(u,\tau/2) \subseteq B_h(u) \subseteq B(u, 3\tau/2), \tag{49}$$

where the inclusions use (20). Applying (47) to radius $\tau/2$ gives

$$2^{-d}d^{-d/2}\tau^d \le \mathrm{Leb}_d(B_h(u)) \le v_d(3\tau/2)^d. \tag{50}$$

This verifies both the mixed-neighborhood mass and the revised anchor; no public-coordinate histogram is used.

Combining the density bounds with these volume bounds yields, for all regimes (with $B_h = B$ under Label LDP),

$$\begin{gathered} c_0\tau^d \le P_X(B(u,\tau)) \le K\tau^d, \qquad c_0\tau^d \le P_X(B_h(u)) \le K\tau^d, \\ 2^{-d}d^{-d/2}\mu \le C_{\mathrm{vol}}(u) \le K\mu \quad (s>0). \end{gathered} \tag{51}$$

For Full LDP the two endpoint displacements change the distance by at most $\sqrt{d}\,h$; for Semi-LDP use (49). Thus in either feature-private regime,

$$B_h(u) \triangle B(u,\tau) \subseteq A_h(u) := \{x \in [0,1]^d : |\|x-u\|_2 - \tau| \le \sqrt{d}\,h\}.$$

Writing $a_+ = \max\{a, 0\}$, polar integration and $h \le \tau$ give

$$\begin{aligned}\mathrm{Leb}_d(A_h(u)) &\le v_d\big[(\tau + \sqrt{d}\,h)^d - (\tau - \sqrt{d}\,h)_+^d\big] \\ &\le 2d^{3/2} v_d h(\tau + \sqrt{d}\,h)^{d-1} \lesssim \tau^{d-1} h. \end{aligned} \tag{52}$$

Keeping x_i fixed and varying the validation location gives

$$Q_X(I_i + I_{i,h} > 0) \lesssim \tau^d, \qquad Q_X(I_i \neq I_{i,h}) \lesssim \tau^{d-1} h \quad (s > 0). \tag{53}$$

For Label LDP the second probability is zero, and the first bound still holds. These are deterministic support bounds conditional on z_i.

For $s > 0$, let $C_{-i,\partial}(u) = \sum_{k \neq i} \mathbf{1}\{x_k \in A_h(u)\}$. Then $|C_{-i,h}(u) - C_{-i}(u)| \le C_{-i,\partial}(u)$ and $|S_{-i,h}(u,r) - S_{-i}(u,r)| \le C_{-i,\partial}(u)$. This binomial variable has mean at most $K\mu h/\tau$ and variance no larger than its mean. Hence

$$\sup_{u,r} \mathbb{E}\big[(C_{-i,h}(u) - C_{-i}(u))^2 + (S_{-i,h}(u,r) - S_{-i}(u,r))^2\big] \lesssim \mu \frac{h}{\tau} + \mu^2 \left(\frac{h}{\tau}\right)^2. \tag{54}$$

Both errors are exactly zero for Label LDP.

Step 2: Raw-noise and stabilized-count errors. At a fixed u let $Z_r = \widetilde{Y}_r(u) - S_{-i,h}(u,r)$. Complete self-report subtraction implies $\mathbb{E}_W[Z_r \mid D] = 0$. Under Full LDP the number of independent Laplace terms in each channel is $(N-1)J_b$. Under Semi-LDP it is

$$K(u) = \sum_{k \neq i} |\mathcal{N}_\tau(u, x_k^{\mathrm{pub}})|.$$

Use $K(u) = (N-1)J_b$ for Full LDP also. In either case

$$\mathbb{E}_W[Z_0^2 + Z_1^2 \mid D] = \frac{16K(u)}{\alpha^2}. \tag{55}$$

For Semi-LDP, a selected private center lies in a coordinate box of side 2τ, so its number is at most $(2\tau/h + 1)^s \le 3^s(\tau/h)^s$. No center can be selected unless $\|x_k^{\mathrm{pub}} - u^{\mathrm{pub}}\|_2 \le \tau$. The public marginal density is at most p_+, without assuming independence between private and public coordinates. Consequently,

$$\begin{aligned}\mathbb{E}K(u) &\le 3^s(\tau/h)^s(N-1)P_X(\|X^{\mathrm{pub}} - u^{\mathrm{pub}}\|_2 \le \tau) \\ &\le 3^s(\tau/h)^s(N-1)p_+(2\tau)^{d-s} \lesssim \mu h^{-s}. \end{aligned} \tag{56}$$

The Full-LDP counterpart follows from (48). For either Label mechanism, $Z_0 = -Z_1$ and

$$\mathbb{E}_W[Z_1^2 \mid D] \le \frac{2C_{-i}(u)}{\alpha^2}, \tag{57}$$

For randomized response, $e^\alpha/(e^\alpha - 1)^2 = 1/[4\sinh^2(\alpha/2)] \le \alpha^{-2}$; the Laplace variance is $2\alpha^{-2}$. Thus, in all regimes,

$$\sup_u \mathbb{E}(Z_0^2 + Z_1^2) \lesssim \frac{\mu}{\alpha^2 h^s}, \tag{58}$$

with $h^0 = 1$ only as a notation convention.

For $s > 0$, let $\widetilde{C}_{-i,h}^{\mathrm{tr}} = \mathrm{trunc}(\widetilde{Y}_0 + \widetilde{Y}_1, 0, N-1)$. Since $C_{-i,h} \in [0, N-1]$, truncation cannot increase its distance from the true count:

$$|\widetilde{C}_{-i,h}^{\mathrm{tr}} - C_{-i,h}| \le |Z_0 + Z_1|. \tag{59}$$

Because $C_{-i,h}$ is an integer, nearest-integer rounding in (37) gives the deterministic inequality

$$|\widetilde{C}_{-i,h} - C_{-i,h}| \le 2|\widetilde{C}_{-i,h}^{\mathrm{tr}} - C_{-i,h}| + 2\omega|C_{\mathrm{vol}} - C_{-i,h}|. \tag{60}$$

Since $C_{-i,h}$ is binomial at fixed u and C_{vol} is deterministic,

$$\mathbb{E}(C_{\mathrm{vol}} - C_{-i,h})^2 = \mathrm{Var}(C_{-i,h}) + (C_{\mathrm{vol}} - \overline{C}_{-i,h})^2 \lesssim \mu + \mu^2.$$

This holds for nonuniform densities: both $\mathrm{Leb}(B_h)$ and $P_X(B_h)$ are $O(\tau^d)$, and equality between them is not required. If $\mu \geq 1$, then

$$\omega^2 \mathbb{E}(C_{\mathrm{vol}} - C_{-i,h})^2 \leq \frac{p_-^2 \mathbb{E}(C_{\mathrm{vol}} - C_{-i,h})^2}{\mu \alpha^2 h^s} \lesssim \frac{\mu}{\alpha^2 h^s}, \qquad \sup_u \mathbb{E}(\widetilde{C}_{-i,h} - C_{-i,h})^2 \lesssim \frac{\mu}{\alpha^2 h^s}. \tag{61}$$

Completing the square in (38), its minimizer is a nearest feasible integer to

$$S_{-i,h}(1) + \frac{\widetilde{C}_{-i,h} - C_{-i,h} + Z_1 - Z_0}{2}.$$

Although $S_{-i,h}(1)$ need not be feasible after the total-count fit, $S_{-i,h}(1) \wedge \widetilde{C}_{-i,h}$ is feasible. Comparing with that point gives

$$\begin{aligned} |\widetilde{S}_{-i,h}(1) - S_{-i,h}(1)| &\leq (S_{-i,h}(1) - \widetilde{C}_{-i,h})_+ + |\widetilde{C}_{-i,h} - C_{-i,h} + Z_1 - Z_0| \\ &\leq 2|\widetilde{C}_{-i,h} - C_{-i,h}| + |Z_1 - Z_0|, \\ |\widetilde{S}_{-i,h}(0) - S_{-i,h}(0)| &\leq 3|\widetilde{C}_{-i,h} - C_{-i,h}| + |Z_1 - Z_0|. \end{aligned}$$

Combining these inequalities with (61) gives

$$\sup_{u,r} \mathbb{E}\big[(\widetilde{C}_{-i,h} - C_{-i,h})^2 + (\widetilde{S}_{-i,h}(r) - S_{-i,h}(r))^2\big] \lesssim \frac{\mu}{\alpha^2 h^s} \qquad (s > 0,\ \mu \geq 1). \tag{62}$$

For Label LDP, $\widetilde{C}_{-i,h} = C_{-i}$ and the noise-free sample label count S_{-i} is feasible. Nearest-feasible-integer optimality yields $|\widetilde{S}_{-i,h}(1) - S_{-i}(u,1)| \leq 2|Z_1|$, with the same error magnitude for label 0. Thus the sharper bound is

$$\mathbb{E}_W[(\widetilde{S}_{-i,h}(r) - S_{-i}(u,r))^2 \mid D] \leq \frac{8C_{-i}(u)}{\alpha^2}. \tag{63}$$

This also implies the common bound with $h^0 = 1$.

Step 3: Small-count events and stabilization. First suppose $\mu < 8/c_0$. Feasibility and (40) give

$$|\Phi_i^\star| \leq \tfrac{1}{2} Q_X(I_i = 1), \qquad |\widetilde{\Phi}_i| \leq \frac{1}{2M} \sum_{t=1}^{M} I_{i,h}(u_t).$$

The indicators in the last sum are i.i.d. conditional on z_i, so (53) implies

$$\mathcal{R}_i \lesssim \tau^{2d} + \frac{\tau^d}{M}.$$

Since $N\tau^d \leq 2\mu < 16/c_0$, this is bounded by a fixed multiple of $(\tau^{2d} + \tau^d/M)e^{-cN\tau^d}$ for any fixed $c > 0$. Thus both claimed bounds hold in this regime. Below assume $\mu \geq 8/c_0$.

At a fixed query define

$$E_0(u) = \{C_{-i}(u) \geq c_0\mu/2,\ C_{-i,h}(u) \geq c_0\mu/2\}, \qquad E(u) = E_0(u) \cap \{\widetilde{C}_{-i,h}(u) \geq c_0\mu/8\}.$$

These are pointwise events, not a uniform-in-query concentration event. The binomial lower-tail inequality and (51) give

$$\sup_u \Pr(E_0(u)^c) \leq 2e^{-c_0\mu/8}. \tag{64}$$

Under Label LDP, $\widetilde{C}_{-i,h} = C_{-i,h} = C_{-i}$, so $E = E_0$ and there is no noise-induced total-count underestimation event.

For $s > 0$, on E_0 we have $C_{-i,h} \geq 4$. If $Z_0 + Z_1 \geq -c_0\mu/8$, then $\widetilde{Y}_0 + \widetilde{Y}_1 \geq C_{-i,h}/2$; truncation at $N - 1$ preserves this lower bound because $C_{-i,h} \leq N - 1$. Thus $\widetilde{C}^{\mathrm{tr}}_{-i,h} \geq C_{-i,h}/2$, and (37) gives $\widetilde{C}_{-i,h} \geq 3C_{-i,h}/8 - 1/2 \geq C_{-i,h}/4 \geq c_0\mu/8$. Therefore

$$E_0(u) \cap E(u)^c \subseteq \{Z_0 + Z_1 < -c_0\mu/8\}. \tag{65}$$

Conditional on D, $Z_0 + Z_1$ is a sum of $2K(u)$ independent $\mathrm{Lap}(2/\alpha)$ variables. For $|t| \le \alpha/4$,

$$\log \mathbb{E}_W[e^{t(Z_0+Z_1)} \mid D] = -2K(u)\log(1 - 4t^2/\alpha^2) \le 16K(u)t^2/\alpha^2.$$

Exponential Markov inequality with $t = \min\{\alpha^2 x/(32K(u)), \alpha/4\}$ implies, if $K(u) > 0$,

$$\Pr_W(Z_0 + Z_1 \le -x \mid D) \le \exp\left[-\min\left\{\frac{\alpha^2 x^2}{64K(u)}, \frac{\alpha x}{8}\right\}\right]. \tag{66}$$

If $K(u) = 0$, the probability is zero for $x > 0$. Under Full LDP, $K(u) \le 3^d \mu h^{-d}$ deterministically. Under Semi-LDP, the number of providers with $\|x_k^{\mathrm{pub}} - u^{\mathrm{pub}}\|_2 \le \tau$ is binomial, with mean at most $p_+ 2^{d-s}(N-1)\tau^{d-s}$. For a binomial variable L, $\log \mathbb{E}e^L \le (e-1)\mathbb{E}L$. Apply this inequality to that public-neighborhood count and use (56) pointwise. For a sufficiently large fixed K, this gives

$$\Pr\{K(u) > K\mu h^{-s}\} \le e^{-c(N-1)\tau^{d-s}} \le e^{-c\mu}. \tag{67}$$

For Full LDP the event is empty after increasing K. Using (66) with $x = c_0\mu/8$ on the complementary event, and $\alpha \le \bar{\alpha}$, proves

$$\sup_u \Pr(E_0(u) \cap E(u)^c) \lesssim e^{-c\mu} + e^{-c\min\{\mu\alpha^2 h^s, \mu\alpha\}} \lesssim e^{-c\mu\alpha^2 h^s}. \tag{68}$$

The reference provides a deterministic magnitude bound, not just a probability bound. From (37) and (51),

$$1 + \widetilde{C}_{-i,h} \ge \omega C_{\mathrm{vol}} + \frac{1}{2} \ge \frac{1+\omega\mu}{2^{d+1}d^{d/2}}.$$

The harmonic envelope and its monotonicity consequently give

$$|\widetilde{\phi}_i(u,r)| \lesssim I_{i,h}(u)\frac{\log(e+\mu)}{\omega\mu}. \tag{69}$$

For example, use $H_m \le 1 + \log m$ and the decreasing function $(1+\log t)/t$ for $t \ge 1$; when $\omega\mu \le 1$, the global bound $|\widetilde{\phi}_i| \le I_{i,h}/2$ already suffices. Finally, $\omega^{-2} = \max\{16, \mu\alpha^2 h^s/p_-^2\}$, so

$$\omega^{-2}\sup_u \Pr(E_0(u) \cap E(u)^c) \lesssim \min\left\{1, \frac{1}{\mu\alpha^2 h^s}\right\}. \tag{70}$$

To see this, put $t = \mu\alpha^2 h^s$ only for this scalar calculation: both $\max\{16, t/p_-^2\}e^{-ct}$ and $t\max\{16, t/p_-^2\}e^{-ct}$ are bounded on $(0,\infty)$, with constants depending on the fixed $p_- > 0$.

Step 4: Risk decomposition and privacy contribution. Conditional on D and the underlying once-generated reports W, $\widetilde{\phi}_i(\cdot)$ is a fixed function. Query-specific broadcasts merely evaluate this function; they do not fit a new function to the validation sample. Therefore

$$\begin{aligned}\mathcal{R}_i &= \mathbb{E}\left(\int(\widetilde{\phi}_i - \phi_i)\,dQ\right)^2 + \frac{1}{M}\mathbb{E}\,\mathrm{Var}_{v\sim Q}(\widetilde{\phi}_i(v) \mid D, W) \\ &\le 2\mathbb{E}\left(\int(\widetilde{\phi}_i - \widehat{\phi}_{i,h})\,dQ\right)^2 + 2\mathbb{E}\left(\int(\widehat{\phi}_{i,h} - \phi_i)\,dQ\right)^2 + \frac{1}{M}\mathbb{E}\int\widetilde{\phi}_i^2\,dQ.\end{aligned} \tag{71}$$

On $E(u)$, Lemma B.1 gives

$$|\widetilde{\phi}_i - \widehat{\phi}_{i,h}|\mathbf{1}_E \lesssim I_{i,h}\frac{\log(e+\mu)}{\mu^2}(|\widetilde{C}_{-i,h} - C_{-i,h}| + |\widetilde{S}_{-i,h}(r) - S_{-i,h}(r)|), \tag{72}$$

$$|\widetilde{\phi}_i|\mathbf{1}_E \lesssim I_{i,h}\frac{\log(e+\mu)}{\mu}. \tag{73}$$

Cauchy–Schwarz over the deterministic support of $I_{i,h}$, followed by (62) and (53), yields for $s > 0$

$$\mathbb{E}\left(\int(\widetilde{\phi}_i - \widehat{\phi}_{i,h})\mathbf{1}_E\,dQ\right)^2 \lesssim \frac{\log^2(e+\mu)\tau^{2d}}{\mu^4}\frac{\mu}{\alpha^2 h^s} \lesssim \frac{\log^2(e+\mu)}{N^2\mu\alpha^2 h^s}. \tag{74}$$

On E_0^c, use $|\widetilde{\phi}_i - \widehat{\phi}_{i,h}| \le I_{i,h}$ and (64), giving a contribution at most $K\tau^{2d}e^{-c\mu}$ to the squared integral. On $E_0 \cap E^c$, the true $C_{-i,h}$ is large, so (69) bounds the difference by $KI_{i,h}\log(e+\mu)/(\omega\mu)$. Cauchy–Schwarz and (70) give

$$\mathbb{E}\left(\int (\widetilde{\phi}_i - \widehat{\phi}_{i,h})\mathbf{1}_{E_0\cap E^c}\, dQ\right)^2 \lesssim \frac{\log^2(e+\mu)}{N^2}\min\left\{1, \frac{1}{\mu\alpha^2 h^s}\right\}.$$

Splitting into the three disjoint events and using $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ proves

$$\mathbb{E}\left(\int (\widetilde{\phi}_i - \widehat{\phi}_{i,h})\, dQ\right)^2 \lesssim \frac{\log^2(e+\mu)}{N^2\mu\alpha^2 h^s} + \tau^{2d}e^{-c\mu} \qquad (s>0). \tag{75}$$

For Label LDP the underestimation event is absent. Using (63) in the same good-event calculation instead gives

$$\mathbb{E}\left(\int (\widetilde{\phi}_i - \phi_i)\, dQ\right)^2 \lesssim \frac{\log^2(e+\mu)}{N^2\mu\alpha^2} + \tau^{2d}e^{-c\mu}. \tag{76}$$

No independence between errors at different validation locations is used.

Step 5: Discretization and validation contributions. For $s > 0$, on E_0 first change the self-indicators at fixed approximate counts, and then change the counts while holding the exact self-indicators. Lemma B.1 gives

$$\begin{aligned}|\widehat{\phi}_{i,h} - \phi_i|\mathbf{1}_{E_0} \lesssim{}& \frac{\log(e+\mu)}{\mu}|I_{i,h} - I_i| \\ &+ I_i\frac{\log(e+\mu)}{\mu^2}(|C_{-i,h}(u) - C_{-i}(u)| + |S_{-i,h}(u,r) - S_{-i}(u,r)|).\end{aligned}$$

The deterministic mismatch bound in (53), followed by Cauchy–Schwarz and (54), gives

$$\begin{aligned}\mathbb{E}\left(\int (\widehat{\phi}_{i,h} - \phi_i)\mathbf{1}_{E_0}\, dQ\right)^2 &\lesssim \frac{\log^2(e+\mu)}{\mu^2}(\tau^{d-1}h)^2 \\ &\quad + \frac{\log^2(e+\mu)\tau^{2d}}{\mu^4}\left[\mu\frac{h}{\tau} + \mu^2\left(\frac{h}{\tau}\right)^2\right] \\ &\lesssim \frac{\log^2(e+\mu)}{N^2}\left[\left(\frac{h}{\tau}\right)^2 + \frac{h}{\mu\tau}\right].\end{aligned} \tag{77}$$

This proof uses the histogram self-indicators $I_{i,h}, R_{i,h}$. If individual validation queries are available to provider i, replacing only these self-indicators by I_i, R_i removes the first term in the pointwise bound above; the same risk upper bound follows, with no self-indicator mismatch contribution. The histogram approximation of the leave-one-out counts remains. On E_0^c, use the amplitude bound on the union of the two supports and (64). Thus the full squared discretization integral is bounded by (77) plus $K\tau^{2d}e^{-c\mu}$. Under Label LDP it is identically zero.

For validation variance, (73) gives

$$\frac{1}{M}\mathbb{E}\int \widetilde{\phi}_i^2\mathbf{1}_E\, dQ \lesssim \frac{\log^2(e+\mu)\tau^d}{M\mu^2}.$$

The E_0^c contribution is at most $K\tau^d e^{-c\mu}/M$. For $s > 0$, (69) and (70) bound the remaining contribution by

$$\frac{K\log^2(e+\mu)\tau^d}{M\mu^2}\min\left\{1, \frac{1}{\mu\alpha^2 h^s}\right\} \le \frac{K\log^2(e+\mu)\tau^d}{M\mu^2}.$$

For Label LDP there is no remaining event. Hence, in every regime,

$$\frac{1}{M}\mathbb{E}\int \widetilde{\phi}_i^2\, dQ \lesssim \frac{\log^2(e+\mu)}{N^2 M\tau^d} + \frac{\tau^d}{M}e^{-c\mu}. \tag{78}$$

Insert these bounds into (71) and use $N/2 \le N-1 < N$, so $\mu \asymp N\tau^d$ and $\log(e+\mu) \asymp L_N$. This proves (42) and (43). □

Proof of Corollary C.2. For $s > 0$, $h_0/\tau \to 0$ and $h_0/2 \le h \le h_0$. Thus the selected bandwidth eventually satisfies both the Full condition $h \le \tau$ and the Semi condition $h \le \tau/\sqrt{s}$. Direct substitution yields

$$\left(\frac{h}{\tau}\right)^2 \asymp \frac{1}{N\alpha^2\tau^d h^s} \asymp (N\alpha^2\tau^{d+s})^{-2/(s+2)}.$$

The boundary-fluctuation term is absorbed because $\mu h/\tau \to \infty$. Indeed, $N\alpha^2\tau^{d+s} \le \bar{\alpha}^2 N\tau^d$, and therefore $\mu h/\tau \gtrsim \mu^{1-1/(s+2)} \to \infty$. Since $M \gtrsim N$ and $\alpha^2 h^s \le \bar{\alpha}^2$,

$$\frac{1}{M\tau^d} \lesssim \frac{1}{N\tau^d} \lesssim \frac{1}{N\alpha^2\tau^d h^s}.$$

The remaining exponential term is negligible under the stated growth condition alone. Indeed, with $T = N\alpha^2\tau^{d+s}$ and $\mu = (N-1)\tau^d$, we have $T \le 2\bar{\alpha}^2\mu$ and $\mu \to \infty$; since $M \gtrsim N$,

$$\left(\tau^{2d} + \frac{\tau^d}{M}\right) e^{-cN\tau^d} \lesssim \frac{\mu^2+\mu}{N^2} e^{-c\mu} = o\left(\frac{L_N^2}{N^2} T^{-2/(s+2)}\right).$$

For Label LDP, use $1/(M\tau^d) \lesssim \bar{\alpha}^2/(N\alpha^2\tau^d)$ in (43). The same remainder argument applies. □

C.2 Proof of the matching lower bounds

Proof of Theorem C.3. Fix $z_i = (x_i, 1)$ with $x_i = (1/2, \dots, 1/2)$. Let Q_0 have uniform features and label 1. Restrict P to distributions with uniform features and $Y \mid X = x \sim \mathrm{Bernoulli}(\eta(x))$, where $1/4 \le \eta(x) \le 3/4$. These distributions belong to the specified classes. Grant the estimator exact knowledge of Q_0, all individual reports, and all designated public features. A finite validation sample can be simulated independently from Q_0, so a lower bound in this enlarged experiment applies to every $M \ge 1$ and to the aggregate-broadcast implementation. Independent estimator randomization can be averaged out under squared loss. Write X for all training features and W for all reports, and retain $\mu = (N-1)\tau^d$. The testing strategy is in the spirit of private minimax lower bounds (Duchi et al., 2018); the channel calculations below are specific to the present protocols.

Step 1: An exact representation of the realized target. For $u \in B(x_i, \tau)$ the retained record has $I_i(u) = R_i(u, 1) = 1$. Using (18) and integrating gives

$$\Phi_i^\star(D) = \int_{B(x_i,\tau)} \frac{H_{C_{-i}(u)+1} - 1/2}{C_{-i}(u)+1}\, du - \sum_{k\neq i} w_k Y_k, \tag{79}$$

$$w_k = \int_{B(x_i,\tau)} \mathbf{1}\{X_k \in B(u,\tau)\} \frac{H_{C_{-i}(u)+1} - 1}{C_{-i}(u)(C_{-i}(u)+1)}\, du. \tag{80}$$

Set the fraction in (80) to zero when $C_{-i}(u) = 0$. Then the representation includes empty neighborhoods. The weights are nonnegative and depend only on X, not on labels.

The ball $B(x_i, \tau/4)$ is untruncated. Its training count has mean $\upsilon_d 4^{-d}\mu$. The count in the cube-truncated ball $B(x_i, 2\tau)$ has mean at most $\upsilon_d 2^d \mu$. A binomial lower-tail bound and exponential Markov inequality at $t = \log 2$ give

$$\Pr\left\{\sum_{k\neq i} \mathbf{1}\{X_k \in B(x_i, \tau/4)\} < \frac{\upsilon_d}{2\cdot 4^d}\mu\right\} \le e^{-\upsilon_d\mu/(8\cdot 4^d)},$$

$$\Pr\left\{\sum_{k\neq i} \mathbf{1}\{X_k \in B(x_i, 2\tau)\} > 2^{d+1}\upsilon_d\mu\right\} \le e^{-(2\log 2 - 1)2^d\upsilon_d\mu}.$$

For sufficiently large μ, with probability at least $1/2$ both counts satisfy the complementary inequalities. On that event, for every $u \in B(x_i, \tau/4)$,

$$B(x_i, \tau/4) \subseteq B(u, \tau) \subseteq B(x_i, 2\tau), \qquad c_d\mu \le C_{-i}(u) \le K_d\mu.$$

If X_k also lies in the inner ball, then $\|X_k - u\|_2 \le \tau/2$ throughout the inner ball. The harmonic lower bound therefore gives

$$w_k \gtrsim \tau^d \frac{\log(e+\mu)}{\mu^2}.$$

There are at least $c_d\mu$ such records on the same event. Taking expectations yields

$$\mathbb{E}_X \sum_{k\neq i} w_k \mathbf{1}\{X_k \in B(x_i, \tau/4)\} \gtrsim \frac{\log(e+\mu)}{N}, \qquad \mathbb{E}_X \sum_{k\neq i} w_k^2 \gtrsim \frac{\log^2(e+\mu)}{N^2\mu}. \tag{81}$$

Step 2: Two alternatives and their report laws. Take

$$\eta_\pm(x) = \frac{1}{2} \pm \delta \mathbf{1}\{x \in B(x_i, \tau/4)\}, \qquad 0 < \delta \leq \frac{1}{4}. \tag{82}$$

Let $P_\pm$ be the training distributions and $\theta_\pm = \mathbb{E}_\pm \Phi_i^\star(D)$. The label-independent term in (79) cancels, so

$$|\theta_+ - \theta_-| = 2\delta \mathbb{E}_X \sum_{k\neq i} w_k \mathbf{1}\{X_k \in B(x_i, \tau/4)\} \gtrsim \delta \frac{\log(e+\mu)}{N}. \tag{83}$$

Write $\mathbb{P}_\pm$ for the observed laws of reports and designated public features, to distinguish them from $P_\pm$.

Full LDP and Semi-LDP Laplace reports. For Semi-LDP condition on the exact public vector $a = x^{\mathrm{pub}}$; omit this conditioning for Full LDP. Let $\pi^\pm_{\ell,r}(a)$ be the conditional probabilities of the private-cell–label one-hot coordinates. With f_0 the product density of unshifted $\mathrm{Lap}(2/\alpha)$ noise, the conditional report laws have densities

$$f_\pm(w \mid a) = f_0(w) \sum_{\ell,r} \pi^\pm_{\ell,r}(a) L_{\ell,r}(w), \qquad L_{\ell,r}(w) = \exp\Big\{ \frac{\alpha}{2}(|w_{\ell,r}| - |w_{\ell,r} - 1|) \Big\}.$$

Under f_0, these likelihood ratios are independent, have mean 1, and lie in $[e^{-\alpha/2}, e^{\alpha/2}]$. Moreover $f_- \geq e^{-\alpha/2} f_0$ and $\sum_{\ell,r}(\pi^+_{\ell,r} - \pi^-_{\ell,r}) = 0$. The KL–chi-square bound and independence imply

$$\begin{aligned}
\mathrm{KL}(f_+, f_-) &\leq e^{\alpha/2} \mathbb{E}_{f_0}\left[\left\{ \sum_{\ell,r} (\pi^+_{\ell,r} - \pi^-_{\ell,r})(L_{\ell,r} - 1) \right\}^2 \right] \\
&= e^{\alpha/2} \sum_{\ell,r} (\pi^+_{\ell,r} - \pi^-_{\ell,r})^2 \mathbb{E}_{f_0}(L_{\ell,r} - 1)^2 \\
&\leq e^{\alpha/2}(e^{\alpha/2} - 1)^2 \sum_{\ell,r} (\pi^+_{\ell,r} - \pi^-_{\ell,r})^2 \lesssim \alpha^2 \sum_{\ell,r} (\pi^+_{\ell,r} - \pi^-_{\ell,r})^2.
\end{aligned} \tag{84}$$

For each private cell let

$$\lambda_\ell(a) = \int_{A^{\mathrm{priv}}_\ell} \mathbf{1}\{(x^{\mathrm{priv}}, a) \in B(x_i, \tau/4)\}\, dx^{\mathrm{priv}}.$$

Uniform features give $\pi^+_{\ell,1} - \pi^-_{\ell,1} = 2\delta\lambda_\ell$ and the negative of this difference for label 0. Since $0 \leq \lambda_\ell \leq h^s$,

$$\sum_{\ell,r} (\pi^+_{\ell,r} - \pi^-_{\ell,r})^2 = 8\delta^2 \sum_\ell \lambda_\ell(a)^2 \leq 8\delta^2 h^s \sum_\ell \lambda_\ell(a), \qquad \int \sum_\ell \lambda_\ell(a)\, da = \upsilon_d 4^{-d} \tau^d.$$

Integrate (84) over the common uniform public-feature law and sum over the $N-1$ independent reports. This gives

$$\mathrm{KL}(\mathbb{P}_+, \mathbb{P}_-) \lesssim \mu\alpha^2\delta^2 h^s \qquad (s > 0). \tag{85}$$

For Full LDP the public integrals are absent and private cells are the full cells, so $s = d$. Exact public coordinates in Semi-LDP have already been included in this information calculation.

Label randomized-response reports. Using (30) and (82),

$$\Pr_\pm(\widetilde{Y} = 1 \mid X = x) = \frac{1}{2} \pm \delta \tanh(\alpha/2) \mathbf{1}\{x \in B(x_i, \tau/4)\}.$$

These Bernoulli parameters lie in $[1/4, 3/4]$. For $a, b \in (0,1)$, $\mathrm{KL}(\mathrm{Bern}(a), \mathrm{Bern}(b)) \le (a-b)^2/[b(1-b)]$. Therefore the one-report conditional divergence is at most

$$\frac{64}{3}\delta^2 \tanh^2(\alpha/2)\mathbf{1}\{x \in B(x_i, \tau/4)\} \le \frac{16}{3}\alpha^2\delta^2\mathbf{1}\{x \in B(x_i, \tau/4)\}.$$

Label Laplace reports. Let $f_0(w) = (\alpha/2)e^{-\alpha|w|}$ and $L(w) = f_0(w-1)/f_0(w)$. Conditional on $X = x$, the report density under $P_\pm$ is $f_\pm = f_0\{1 - \eta_\pm(x) + \eta_\pm(x)L\}$. Since $e^{-\alpha} \le L \le e^\alpha$, we have $f_- \ge e^{-\alpha} f_0$ and

$$\begin{aligned}\mathrm{KL}(f_+, f_-) &\le e^\alpha\{\eta_+(x) - \eta_-(x)\}^2 \mathbb{E}_{f_0}(L-1)^2 \\ &\le 4e^\alpha(e^\alpha - 1)^2\delta^2\mathbf{1}\{x \in B(x_i, \tau/4)\} \\ &\lesssim \alpha^2\delta^2\mathbf{1}\{x \in B(x_i, \tau/4)\}.\end{aligned}$$

The last inequality uses $\alpha \le \bar{\alpha}$. Integrating either Label divergence bound over X and summing independent reports yields

$$\mathrm{KL}(\mathbb{P}_+, \mathbb{P}_-) \lesssim \mu\alpha^2\delta^2 \qquad (s = 0). \tag{86}$$

The fixed provider's report has the same law under both hypotheses and adds no divergence. Choose

$$\delta = c_* \min\{1, (\mu\alpha^2 h^s)^{-1/2}\},$$

with $h^0 = 1$ and a sufficiently small fixed $0 < c_* \le 1/4$. Equations (85)–(86) make the total KL at most $1/8$, so Pinsker's inequality gives

$$\mathrm{TV}(\mathbb{P}_+, \mathbb{P}_-) \le \frac{1}{4}. \tag{87}$$

Step 3: A testing reduction for a dataset-dependent target. Given an estimator T, select the closer of θ_+ and θ_-, with deterministic tie breaking. On a testing error under hypothesis ν, $|T - \theta_\nu| \ge |\theta_+ - \theta_-|/2$. The sum of testing errors is at least $1 - \mathrm{TV}(\mathbb{P}_+, \mathbb{P}_-)$, so

$$\sum_{\nu \in \{+,-\}} \mathbb{E}_\nu(T - \theta_\nu)^2 \ge \frac{(\theta_+ - \theta_-)^2}{4}[1 - \mathrm{TV}(\mathbb{P}_+, \mathbb{P}_-)].$$

Since the target is the realized $\Phi_i^\star(D)$,

$$\mathbb{E}_\nu(T - \theta_\nu)^2 \le 2\mathbb{E}_\nu(T - \Phi_i^\star)^2 + 2\,\mathrm{Var}_\nu(\Phi_i^\star).$$

Consequently,

$$\max_{\nu \in \{+,-\}} \mathbb{E}_\nu(T - \Phi_i^\star)^2 \ge \frac{(\theta_+ - \theta_-)^2}{16}[1 - \mathrm{TV}(\mathbb{P}_+, \mathbb{P}_-)] - \max_\nu \mathrm{Var}_\nu(\Phi_i^\star). \tag{88}$$

We now control the variance term. Replace z_k, $k \ne i$, by an independent copy z_k' to form $D^{(k)}$. At a fixed u let $C_{-\{i,k\}}(u) = \sum_{\ell \notin \{i,k\}} \mathbf{1}\{X_\ell \in B(u, \tau)\}$. For $N \ge 3$, its mean is at least $c_0\mu/2$, so

$$\Pr\{C_{-\{i,k\}}(u) < c_0\mu/4\} \le e^{-c\mu}.$$

Replacing z_k changes each of C_{-i}, S_{-i} by at most 1, and makes no change unless X_k or X_k' lies in $B(u, \tau)$. The latter event has probability at most $K\tau^d$ and is independent of $C_{-\{i,k\}}(u)$. On the large-count event use Lemma B.1; on its complement use the global amplitude bound. Uniformly for $u \in B(x_i, \tau)$ this gives

$$\mathbb{E}\Big[(\phi_i(u, 1; D) - \phi_i(u, 1; D^{(k)}))^2\Big] \lesssim \tau^d\left[\frac{\log^2(e+\mu)}{\mu^4} + e^{-c\mu}\right].$$

Efron–Stein and Cauchy–Schwarz over the deterministic validation support now imply

$$\begin{aligned}\mathrm{Var}(\Phi_i^\star) &\le \frac{1}{2}\sum_{k \ne i} \mathbb{E}\Big[(\Phi_i^\star(D) - \Phi_i^\star(D^{(k)}))^2\Big] \\ &\lesssim N\tau^{3d}\left[\frac{\log^2(e+\mu)}{\mu^4} + e^{-c\mu}\right] \lesssim \frac{\log^2(e+\mu)}{N^2\mu}.\end{aligned} \tag{89}$$

The last step uses $N\tau^d \asymp \mu$ and boundedness of $\mu^4 e^{-c\mu}/\log^2(e+\mu)$ for $\mu \ge 1$. Combining (83), (87), and (88) gives fixed $c_1, c_2 > 0$ with

$$\max_\nu \mathbb{E}_\nu(T - \Phi_i^\star)^2 \ge c_1\frac{\log^2(e+\mu)}{N^2}\min\left\{1, \frac{1}{\mu\alpha^2 h^s}\right\} - c_2\frac{\log^2(e+\mu)}{N^2\mu}. \tag{90}$$

Step 4: Removing the target-variance subtraction. Take the auxiliary distribution P_0 with $\eta \equiv 1/2$. Only for this auxiliary bound, also reveal all training features X. This makes estimation easier and does not change the observation model used in the KL calculations above. Conditional on X, the labels are independent fair Bernoulli variables. For the Laplace reports, changing only a label gives likelihood ratio in $[e^{-\alpha}, e^{\alpha}]$. For RR the two nontrivial ratios are exactly e^{α} and $e^{-\alpha}$. Hence Bayes' rule gives, for all specified reports,

$$\mathrm{Var}_0(Y_k \mid X, W) \geq \frac{e^{\alpha}}{(1+e^{\alpha})^2} \geq \frac{e^{\bar{\alpha}}}{(1+e^{\bar{\alpha}})^2} > 0. \tag{91}$$

The posterior factors over providers because both the prior and the reporting likelihood factor. The weights in (79) are fixed given X. Therefore, for any estimator measurable with respect to X, W,

$$\mathbb{E}_0[(T-\Phi_i^\star)^2 \mid X, W] \geq \mathrm{Var}_0(\Phi_i^\star \mid X, W) = \sum_{k \neq i} w_k^2 \,\mathrm{Var}_0(Y_k \mid X, W). \tag{92}$$

This also applies to estimators using less information. Taking expectations and using (81) gives a fixed $c_3 > 0$ such that

$$\mathbb{E}_0(T-\Phi_i^\star)^2 \geq c_3 \frac{\log^2(e+\mu)}{N^2\mu}. \tag{93}$$

For the same arbitrary estimator, its supremum risk dominates both (90) and (93). Multiply the second by c_2/c_3 and add:

$$\begin{aligned}\left(1+\frac{c_2}{c_3}\right) \sup_{P \in \mathcal{P}} \mathbb{E}_{P,Q_0}[(T-\Phi_i^\star)^2 \mid z_i] &\geq \max_{\nu} \mathbb{E}_\nu (T-\Phi_i^\star)^2 + \frac{c_2}{c_3}\mathbb{E}_0(T-\Phi_i^\star)^2 \\ &\geq c_1 \frac{\log^2(e+\mu)}{N^2} \min\left\{1, \frac{1}{\mu\alpha^2 h^s}\right\}.\end{aligned}$$

All constants may depend on $\bar{\alpha}$ but not on the particular $\alpha \in (0, \bar{\alpha}]$. Taking the infimum over estimators, including Q_0 in the supremum, and using $\mu \asymp N\tau^d$ proves (45).

Step 5: Substitution of the selected resolution. For $s > 0$, the grid of Corollary C.2 satisfies $h \asymp h_0$ eventually, and

$$(\mu\alpha^2 h^s)^{-1} \asymp (N\alpha^2\tau^{d+s})^{-2/(s+2)} \longrightarrow 0.$$

For $s = 0$ the same equality in order holds with $h^0 = 1$. The minimum in (45) therefore selects its second argument for sufficiently large N, proving (46). □

Remark 1 (Scope of the rate statement). The rates compare mean-squared estimation of the same realized-dataset TKNN functional under the specified reporting experiments. They do not imply superiority over KNN-Shapley on ranking recovery, corruption AUROC, or downstream prediction. The Label upper bound retains the exact RR variance factor, but the matching statement uses bounded $\alpha \leq \bar{\alpha}$ and constants depending on $\bar{\alpha}$; it is not a uniform sharp characterization when $\alpha \to \infty$. The matching claim concerns the leading selected-resolution rate for $M \gtrsim N$; no matching lower bound for the separate finite-M validation term is asserted.

C.3 From valuation error to ranking AUC

Fix the realized candidate pool D and an external binary marker a_i (e.g., planted corruption) specified before reporting, with nonempty index sets $I_1 = \{i : a_i = 1\}$ and $I_0 = \{i : a_i = 0\}$. For detection by low value, define the exact and estimated scores $S_i = -N\Phi_i^\star$ and $\widetilde{S}_i = -N\widetilde{\Phi}_i$; multiplication by the positive factor N preserves the low-value ranking induced by the negative valuations. For any score vector s, let

$$\mathrm{AUC}_{D,a}(s) = \frac{1}{|I_1||I_0|} \sum_{i \in I_1, j \in I_0} \psi(s_i - s_j), \qquad \psi(u) = \mathbf{1}\{u > 0\} + \tfrac{1}{2}\mathbf{1}\{u = 0\}. \tag{94}$$

For $t > 0$, define the small-gap fraction $\mu_{D,a}(t) = (|I_1||I_0|)^{-1} \sum_{i \in I_1, j \in I_0} \mathbf{1}\{|S_i - S_j| \leq t\}$ and the pairwise estimation risk

$$\mathcal{E}_{D,a} = \frac{1}{|I_1||I_0|} \sum_{i \in I_1, j \in I_0} \mathbb{E}\Big[\big\{(\widetilde{S}_i - \widetilde{S}_j) - (S_i - S_j)\big\}^2 \mid D, a\Big]. \tag{95}$$

The expectation here is over reporting and validation randomness, conditional on the same realized pool and markers.

Theorem C.4 (Fixed-pool AUC stability). *For any fixed D, a as above and every $t > 0$,*

$$\mathbb{E}\Big[\Big|\mathrm{AUC}_{D,a}(\widetilde{S}) - \mathrm{AUC}_{D,a}(S)\Big| \mid D, a\Big] \le \mu_{D,a}(t) + \frac{\mathcal{E}_{D,a}}{t^2}. \tag{96}$$

If D, a are random, $\mathbb{E}\mu_{D,a}(t) \le At^\beta$ for $0 < t \le t_0$, and $\mathbb{E}\mathcal{E}_{D,a} \le r_N$, where $A, \beta, t_0 > 0$ and $r_N \ge 0$, then for $r_N \le At_0^{\beta+2}$,

$$\mathbb{E}\Big|\mathrm{AUC}_{D,a}(\widetilde{S}) - \mathrm{AUC}_{D,a}(S)\Big| \le 2A^{2/(\beta+2)} r_N^{\beta/(\beta+2)}. \tag{97}$$

The small-gap term measures how often exact scores nearly tie across groups; mean-squared error alone cannot control their ordering. The theorem compares against exact TKNN valuation of the *same* pool, not the Bayes-optimal corruption detector.

A sufficient input for the pairwise risk is a lower bound on both group sizes. If $|I_0|, |I_1| \ge \kappa N$ almost surely for some $\kappa > 0$ and $\sup_i \mathbb{E}(\widetilde{\Phi}_i - \Phi_i^\star)^2 \le b_N$, then

$$\mathbb{E}\mathcal{E}_{D,a} \le \frac{2N^2 b_N}{\kappa}. \tag{98}$$

Thus (97) applies with $r_N = 2N^2 b_N/\kappa$ when its small-gap and threshold conditions also hold. Substituting the individual bounds above for b_N requires their i.i.d. sampling assumptions to hold for the marked pool; the planted-corruption experiment need not satisfy those assumptions. Neither the small-gap condition nor separation of corrupt and clean records follows from the valuation MSE bound. Releasing locally retained scores also requires separate privacy analysis.

Proof of Theorem C.4. Fix D, a, and write $\Delta_{ij} = S_i - S_j$ and $\widetilde{\Delta}_{ij} = \widetilde{S}_i - \widetilde{S}_j$ for a cross-group pair. If $|\Delta_{ij}| > t$ and $|\widetilde{\Delta}_{ij} - \Delta_{ij}| < t$, the two gaps have the same strict sign, so $\psi(\widetilde{\Delta}_{ij}) = \psi(\Delta_{ij})$. Since $0 \le \psi \le 1$, for every pair,

$$|\psi(\widetilde{\Delta}_{ij}) - \psi(\Delta_{ij})| \le \mathbf{1}\{|\Delta_{ij}| \le t\} + \mathbf{1}\{|\widetilde{\Delta}_{ij} - \Delta_{ij}| \ge t\}.$$

Take the cross-group average, use the triangle inequality, and then take conditional expectations. The pointwise bound $\mathbf{1}\{|x| \ge t\} \le x^2/t^2$ gives (96). No independence among providers' score errors is used.

For completeness, with $\rho_i(D, a) = \mathbb{E}[(\widetilde{S}_i - S_i)^2 \mid D, a]$, $(x - y)^2 \le 2x^2 + 2y^2$ also yields

$$\mathcal{E}_{D,a} \le 2\left\{\frac{1}{|I_1|}\sum_{i \in I_1} \rho_i(D, a) + \frac{1}{|I_0|}\sum_{j \in I_0} \rho_j(D, a)\right\}.$$

If both group sizes are at least κN, their disjoint union is the candidate pool, so

$$\mathcal{E}_{D,a} \le \frac{2}{\kappa N}\sum_{i=1}^{N} \rho_i(D, a).$$

Taking expectations and using $\mathbb{E}\rho_i = N^2\mathbb{E}(\widetilde{\Phi}_i - \Phi_i^\star)^2$ proves (98) under the stated uniform bound b_N. This step does not assume independent score errors across providers. For the second assertion, average (96) over D, a. When $r_N > 0$, choose $t = (r_N/A)^{1/(\beta+2)} \le t_0$; the two resulting terms are each $A^{2/(\beta+2)} r_N^{\beta/(\beta+2)}$. If $r_N = 0$, every pairwise gap is estimated exactly almost surely and the AUC difference is zero. This proves (97). □

D Details for Real-Data Experiments

D.1 Protocol and Reporting

Data and splits. We use the nine binary datasets in Table 1. Each of 12 repeated splits contains 200 public, 800 training, 80 validation, and 500 test records; training and validation classes are balanced. We draw ten independent privacy reports per split.

Preprocessing. Features are min–max scaled to $[0, 1]$ using the public data.

Parameters and reports. We use Euclidean distance and $k = 5$. Five-fold cross-validation on the public data selects $c_\star \in \{.20, .30, .40, .50, .65, .80, 1, 1.20, 1.40, 1.60, 2, 2.50\}$ by TKNN balanced accuracy, with $\tau = 1.25 c_\star L_{\mathrm{pub}} N^{-1/(4d)}$ and L_{pub} the median public distance. The selected threshold is fixed across privacy budgets within each split. For semi-LDP, the first $s = \min\{\lfloor d/2 \rfloor + 2, d - 1\}$ features are private. Under full LDP and semi-LDP, PIVOTS uses grid size $q = \max\{2, \lceil h_0^{-1} \rceil, \lceil \tau^{-1} \rceil\}$ with h_0 from Corollary C.2. Appendix A specifies the PIVOTS reporting mechanisms.

LDP-KNN baseline. Under full LDP and semi-LDP, LDP-KNN allocates half of α to the private features and half to the label. It adds independent Laplace noise of scale $2 \sum_{j \le s} (b_j - a_j)/\alpha$ to each private feature, where $[a_j, b_j]$ is its public range, and scale $2/\alpha$ to the label. It then clips noisy features to their public bounds and thresholds the noisy label at $1/2$ before computing KNN-Shapley. Under label LDP, features and distances are exact; both methods use binary randomized response with flip probability $r = (1 + e^\alpha)^{-1}$ and debias the reported bit as $(B - r)/(1 - 2r)$.

Corruption and metrics. We corrupt 25% of training records by either flipping 100 labels per class or adding independent Gaussian feature noise to 200 records, scaled by the public feature means and clipped to $[0, 1]$. Recovery is top-80 overlap with each method's own non-private ranking on the same corrupted pool (KNN-Shapley for LDP-KNN; TKNN-Shapley for PIVOTS). Detection is AUROC of negative valuation against the corruption mask. For downstream prediction, we fit logistic regression ($\ell_2 = .001$) on the top 400 selected records and evaluate balanced accuracy on clean test data.

Reporting. Tables report mean $\pm$ split SD after averaging the ten privacy runs within each split. Bold marks the higher private-method mean; $*$ marks paired two-sided Wilcoxon $p < .05$ after Holm correction across datasets within each comparison. Random and NP-KNN are reference baselines. These tests are exploratory.

D.2 RECOVERY

D.2.1 FULL LDP

Table 2: Full LDP: recovery (%). Mean ± split SD.

		Mislabel (25%)		Feature corruption (25%)	
Dataset	α	LDP-KNN	PIVOTS	LDP-KNN	PIVOTS
CODRNA	1	10.69 ±5.4	**21.35** ±6.4*	8.93 ±6.7	**22.81** ±8.2*
	2	11.31 ±5.5	**21.49** ±5.8*	8.33 ±5.8	**22.80** ±7.4*
	3	11.09 ±5.2	**21.73** ±5.7*	8.49 ±5.5	**23.26** ±7.1*
	4	11.50 ±5.6	**21.89** ±6.0*	8.54 ±6.0	**23.55** ±7.0*
SVMG1	1	13.32 ±7.7	**21.76** ±4.1*	1.96 ±2.2	**20.20** ±4.4*
	2	14.82 ±7.3	**21.39** ±4.0*	2.52 ±3.1	**20.27** ±5.5*
	3	16.29 ±7.6	**21.70** ±3.8	3.43 ±3.8	**20.75** ±4.9*
	4	17.12 ±7.8	**22.95** ±3.5	3.77 ±3.9	**22.25** ±3.9*
MAGIC	1	8.70 ±4.9	**19.74** ±6.7*	1.61 ±1.4	**20.04** ±6.6*
	2	8.71 ±4.9	**19.78** ±6.7*	1.66 ±1.8	**20.33** ±5.8*
	3	8.58 ±4.8	**20.22** ±6.0*	1.54 ±1.7	**20.29** ±4.2*
	4	8.35 ±5.3	**21.19** ±5.8*	1.47 ±1.5	**21.15** ±4.0*
SKIN	1	14.84 ±7.5	**19.17** ±7.4	4.29 ±2.8	**14.93** ±5.1*
	2	13.88 ±8.9	**19.37** ±7.0	3.96 ±3.1	**13.98** ±5.8*
	3	18.20 ±8.8	**20.14** ±6.4	3.96 ±4.0	**18.40** ±8.4*
	4	20.97 ±7.7	**22.16** ±6.4	6.67 ±6.4	**22.78** ±7.9*
PHON	1	17.28 ±7.0	**26.17** ±4.0*	9.71 ±5.8	**20.41** ±7.0*
	2	18.10 ±7.9	**26.13** ±4.0*	9.94 ±6.3	**20.44** ±7.4*
	3	18.25 ±7.4	**27.24** ±3.9*	9.98 ±6.9	**20.95** ±7.5*
	4	18.32 ±7.5	**27.07** ±4.0*	10.44 ±6.7	**21.45** ±7.1*
WATER	1	4.73 ±2.0	**23.30** ±24.0*	1.39 ±1.1	**25.58** ±23.7*
	2	4.71 ±2.1	**23.50** ±23.9*	1.52 ±1.0	**25.63** ±23.8*
	3	4.89 ±2.0	**23.53** ±23.9*	1.61 ±1.3	**25.78** ±23.7*
	4	4.52 ±2.0	**23.76** ±23.8*	1.92 ±1.4	**25.89** ±23.7*
CALIF	1	11.05 ±5.4	**23.30** ±5.0*	5.51 ±4.5	**26.81** ±5.5*
	2	11.27 ±4.9	**24.17** ±5.0*	5.55 ±4.4	**27.28** ±5.5*
	3	11.48 ±4.9	**24.76** ±5.4*	5.62 ±4.3	**27.72** ±5.5*
	4	11.77 ±5.2	**26.06** ±5.6*	6.08 ±4.5	**28.98** ±5.8*
HOUSE8L	1	5.06 ±2.7	**20.61** ±4.2*	0.76 ±0.9	**24.31** ±4.8*
	2	5.17 ±2.9	**22.30** ±4.6*	0.84 ±0.8	**23.51** ±4.6*
	3	5.47 ±2.9	**23.62** ±4.7*	0.94 ±1.0	**23.82** ±4.2*
	4	5.46 ±2.7	**24.30** ±4.9*	1.02 ±1.3	**22.85** ±4.2*
PUMP	1	4.72 ±5.0	**19.65** ±5.2*	0.16 ±0.3	**22.45** ±5.0*
	2	4.74 ±5.1	**19.72** ±4.0*	0.20 ±0.4	**21.72** ±4.9*
	3	4.75 ±4.9	**19.66** ±3.5*	0.31 ±0.4	**20.56** ±5.7*
	4	4.67 ±4.9	**21.54** ±2.7*	0.30 ±0.5	**20.33** ±5.5*

D.2.2 SEMI-LDP

Table 3: Semi-LDP: recovery (%). Mean ± split SD.

		Mislabel (25%)		Feature corruption (25%)	
Dataset	α	LDP-KNN	PIVOTS	LDP-KNN	PIVOTS
CODRNA	1	15.65 ±7.6	**28.77** ±10.0*	14.51 ±10.2	**29.56** ±9.5*
	2	15.14 ±7.2	**28.71** ±9.7*	15.06 ±9.7	**29.32** ±9.7*
	3	15.28 ±6.9	**28.61** ±9.3*	15.49 ±10.0	**29.46** ±9.6*
	4	15.40 ±6.9	**28.18** ±9.0*	16.60 ±10.2	**29.51** ±9.6*
SVMG1	1	20.22 ±8.2	**30.71** ±4.3*	6.56 ±5.3	**25.82** ±2.2*
	2	21.08 ±8.0	**30.89** ±4.0*	6.35 ±5.4	**25.65** ±3.1*
	3	21.60 ±7.8	**30.98** ±4.6*	6.48 ±5.8	**27.06** ±3.1*
	4	21.61 ±8.0	**31.66** ±4.9*	6.42 ±6.0	**28.08** ±4.9*
MAGIC	1	9.59 ±4.8	**28.81** ±8.0*	2.24 ±1.6	**28.66** ±7.0*
	2	10.24 ±4.8	**28.25** ±7.2*	2.53 ±1.6	**29.11** ±6.4*
	3	11.09 ±4.9	**27.66** ±6.7*	2.55 ±1.7	**28.19** ±6.4*
	4	11.47 ±5.2	**27.85** ±6.0*	2.59 ±2.0	**29.26** ±6.2*
SKIN	1	15.43 ±8.5	**22.82** ±5.7*	3.39 ±3.0	**19.90** ±8.5*
	2	19.83 ±9.6	**26.49** ±4.9*	5.90 ±5.9	**24.82** ±6.1*
	3	20.57 ±8.3	**35.25** ±8.3*	7.93 ±8.0	**28.75** ±6.8*
	4	23.81 ±9.9	**34.80** ±7.7*	9.69 ±10.2	**31.67** ±7.4*
PHON	1	18.64 ±8.3	**33.84** ±6.2*	12.60 ±6.7	**26.97** ±5.4*
	2	19.17 ±8.5	**34.28** ±5.8*	14.75 ±7.8	**26.60** ±4.7*
	3	19.30 ±8.9	**34.82** ±5.7*	15.68 ±8.9	**26.78** ±4.4*
	4	20.67 ±8.5	**34.97** ±5.7*	17.51 ±9.4	**26.96** ±4.4*
WATER	1	5.85 ±1.8	**30.33** ±22.4*	6.22 ±2.2	**35.25** ±21.3*
	2	6.04 ±1.8	**30.30** ±22.3*	6.52 ±2.5	**35.41** ±21.3*
	3	6.11 ±2.0	**30.40** ±22.3*	6.68 ±2.2	**35.64** ±21.1*
	4	6.05 ±2.3	**30.49** ±22.4*	6.69 ±2.6	**35.58** ±21.1*
CALIF	1	14.70 ±3.8	**29.09** ±6.9*	11.72 ±5.3	**31.33** ±4.2*
	2	14.81 ±4.0	**28.94** ±6.9*	13.35 ±6.2	**31.26** ±4.1*
	3	15.84 ±3.2	**28.97** ±6.9*	13.16 ±6.3	**31.35** ±4.0*
	4	15.84 ±3.3	**29.06** ±6.9*	14.09 ±6.5	**31.44** ±4.1*
HOUSE8L	1	9.65 ±4.8	**32.58** ±6.3*	4.17 ±2.2	**37.03** ±7.8*
	2	10.19 ±5.0	**32.68** ±6.0*	4.28 ±2.1	**36.65** ±7.4*
	3	10.51 ±4.9	**32.80** ±5.9*	4.21 ±2.3	**37.02** ±6.9*
	4	10.88 ±4.8	**33.38** ±5.6*	4.22 ±2.7	**37.29** ±6.9*
PUMP	1	5.79 ±5.2	**37.00** ±6.0*	0.68 ±1.1	**38.38** ±4.2*
	2	5.91 ±5.4	**36.91** ±6.0*	0.49 ±0.8	**38.32** ±4.8*
	3	5.99 ±5.3	**36.79** ±6.0*	0.40 ±0.6	**38.07** ±5.1*
	4	5.95 ±5.0	**36.86** ±5.8*	0.39 ±0.5	**37.59** ±5.1*

D.2.3 LABEL LDP

Table 4: Label LDP: recovery (%). Mean ± split SD.

		Mislabel (25%)		Feature corruption (25%)	
Dataset	α	LDP-KNN	PIVOTS	LDP-KNN	PIVOTS
CODRNA	1	66.53 ±2.0	**78.50** ±4.5*	62.21 ±4.0	**77.83** ±3.5*
	2	82.65 ±2.4	**90.11** ±3.2*	78.58 ±1.8	**88.32** ±2.6*
	3	90.57 ±1.5	**94.42** ±1.7*	87.47 ±1.7	**93.65** ±2.2*
	4	95.09 ±1.3	**97.19** ±0.9*	93.56 ±1.5	**96.52** ±1.1*
SVMG1	1	60.93 ±3.7	**72.05** ±4.8*	45.61 ±7.5	**65.88** ±5.5*
	2	78.34 ±2.6	**84.54** ±3.5*	64.18 ±7.5	**80.86** ±3.7*
	3	87.99 ±1.8	**90.67** ±2.9*	77.91 ±5.5	**88.92** ±2.6*
	4	93.58 ±1.4	**94.44** ±2.2	86.56 ±5.0	**92.96** ±2.0*
MAGIC	1	67.40 ±3.4	**86.45** ±3.5*	62.68 ±6.5	**83.70** ±3.2*
	2	83.41 ±2.6	**93.88** ±1.0*	79.79 ±5.1	**91.52** ±2.3*
	3	90.61 ±1.7	**96.55** ±0.9*	88.86 ±3.3	**95.52** ±1.9*
	4	94.60 ±1.4	**98.08** ±0.6*	93.85 ±2.4	**97.45** ±0.7*
SKIN	1	57.60 ±5.1	**63.56** ±8.3	29.47 ±6.1	**55.30** ±8.5*
	2	76.81 ±3.8	**80.68** ±6.9	41.01 ±8.1	**70.59** ±9.6*
	3	85.71 ±3.2	**88.35** ±5.7	50.18 ±7.3	**82.14** ±9.5*
	4	92.66 ±2.2	**93.09** ±3.9	64.23 ±6.5	**87.02** ±7.6*
PHON	1	64.95 ±3.6	**75.65** ±5.1*	57.48 ±3.8	**70.62** ±5.3*
	2	81.05 ±3.3	**87.78** ±3.3*	76.21 ±4.4	**84.34** ±3.3*
	3	90.14 ±1.4	**92.67** ±2.5*	86.06 ±3.4	**90.38** ±2.2*
	4	95.02 ±0.8	**96.09** ±1.9	92.43 ±2.5	**94.16** ±2.0
WATER	1	73.32 ±2.4	**85.59** ±5.1*	76.05 ±2.3	**86.80** ±4.3*
	2	87.26 ±1.9	**93.71** ±2.3*	88.73 ±1.8	**93.90** ±2.2*
	3	93.46 ±1.3	**96.28** ±1.5*	94.17 ±1.4	**96.74** ±1.3*
	4	96.34 ±0.7	**97.81** ±0.9*	96.86 ±1.0	**98.18** ±0.9*
CALIF	1	71.15 ±3.2	**84.06** ±4.2*	68.54 ±5.0	**81.88** ±4.8*
	2	85.79 ±3.5	**91.89** ±2.9*	84.22 ±3.2	**90.63** ±3.0*
	3	92.48 ±2.0	**95.78** ±1.6*	91.59 ±1.6	**94.50** ±2.1*
	4	95.96 ±1.4	**97.59** ±1.2	95.26 ±1.0	**96.91** ±1.4*
HOUSE8L	1	68.89 ±3.8	**81.88** ±3.7*	62.19 ±5.2	**79.16** ±5.4*
	2	84.42 ±3.2	**90.56** ±2.8*	79.17 ±4.6	**89.08** ±3.7*
	3	91.58 ±1.9	**94.71** ±1.7*	87.81 ±3.9	**93.45** ±2.7*
	4	95.44 ±1.1	**96.90** ±1.2*	93.31 ±2.6	**95.95** ±2.2*
PUMP	1	64.57 ±3.2	**81.38** ±4.4*	55.69 ±5.9	**76.54** ±6.1*
	2	80.70 ±2.7	**90.92** ±2.7*	72.85 ±4.9	**87.99** ±2.9*
	3	89.16 ±2.2	**95.23** ±1.8*	84.16 ±3.1	**92.76** ±1.8*
	4	94.24 ±1.3	**97.14** ±1.5*	91.24 ±2.6	**95.44** ±1.4*

D.3 DETECTION

Both methods are evaluated against the same planted corruption mask; higher AUROC means corrupted records tend to receive lower values.

D.3.1 FULL LDP

Table 5: Full LDP: detection AUROC (%). Mean ± split SD. Random is shared across budgets.

		Mislabel (25%)			Feature corruption (25%)		
Dataset	α	Random	LDP-KNN	PIVOTS	Random	LDP-KNN	PIVOTS
CODRNA	1	50.08 ±0.5	62.22 ±4.5	**67.72** ±4.3*	50.13 ±0.6	51.30 ±6.6	**62.49** ±2.8*
	2		63.03 ±4.7	**68.26** ±4.2*		49.27 ±6.5	**62.52** ±2.8*
	3		63.15 ±4.7	**68.60** ±4.2*		46.76 ±7.4	**62.64** ±2.8*
	4		63.22 ±4.7	**68.93** ±4.1*		44.58 ±7.7	**62.62** ±2.8*
SVMG1	1	49.85 ±0.6	85.11 ±4.3	**91.58** ±2.0*	49.89 ±0.6	59.35 ±2.9	**75.46** ±2.5*
	2		86.69 ±4.0	**91.90** ±1.8*		60.47 ±2.8	**75.39** ±2.3*
	3		87.48 ±4.0	**92.16** ±1.7*		61.77 ±3.2	**75.22** ±2.3*
	4		88.14 ±3.9	**92.29** ±1.7*		62.50 ±3.3	**74.98** ±2.4*
MAGIC	1	50.05 ±1.0	61.42 ±3.7	**64.47** ±2.2*	50.06 ±0.7	42.30 ±3.6	**54.81** ±2.7*
	2		62.12 ±3.9	**65.71** ±2.8		42.11 ±3.6	**54.51** ±2.5*
	3		62.84 ±4.0	**66.56** ±3.0		42.13 ±3.9	**54.17** ±2.3*
	4		63.43 ±4.2	**67.37** ±3.0		42.62 ±4.1	**53.95** ±2.2*
SKIN	1	50.47 ±0.5	93.67 ±1.7	**96.95** ±2.4*	49.88 ±0.5	67.93 ±5.8	**85.03** ±7.1*
	2		96.30 ±0.5	**97.47** ±2.0		78.58 ±3.7	**84.40** ±8.0
	3		97.47 ±0.4	**98.09** ±1.2		85.08 ±2.4	**85.98** ±5.3
	4		98.18 ±0.4	**98.62** ±0.6		**88.43** ±1.7	87.13 ±3.8
PHON	1	50.24 ±0.8	72.69 ±4.7	**74.28** ±3.8	49.89 ±0.5	47.36 ±2.1	**60.53** ±3.5*
	2		74.48 ±3.6	**74.76** ±3.9		47.86 ±2.4	**60.56** ±3.6*
	3		**75.82** ±2.6	75.15 ±3.8		48.30 ±2.6	**60.57** ±3.8*
	4		**76.84** ±2.1	75.49 ±3.8		48.87 ±2.9	**60.50** ±3.7*
WATER	1	50.15 ±0.5	47.75 ±2.2	**51.56** ±1.7*	50.08 ±0.6	50.12 ±4.6	**50.51** ±1.3
	2		47.82 ±2.0	**51.62** ±1.7*		48.21 ±4.6	**50.50** ±1.4
	3		47.76 ±2.0	**51.67** ±1.8*		48.34 ±6.4	**50.49** ±1.5
	4		47.72 ±1.9	**51.73** ±1.8*		49.28 ±6.8	**50.48** ±1.5
CALIF	1	49.92 ±0.6	62.22 ±4.9	**67.78** ±4.3*	49.82 ±0.7	50.24 ±4.9	**63.49** ±3.0*
	2		62.45 ±5.0	**68.11** ±4.3		50.63 ±3.4	**63.30** ±2.7*
	3		62.67 ±4.9	**68.50** ±4.5		51.08 ±3.0	**62.74** ±2.6*
	4		62.80 ±5.0	**68.96** ±4.6		50.50 ±4.5	**62.24** ±2.4*
HOUSE8L	1	49.85 ±0.7	63.80 ±4.6	**68.24** ±3.5	50.05 ±0.5	47.50 ±3.7	**62.64** ±3.1*
	2		64.02 ±4.4	**69.42** ±3.6		46.52 ±3.9	**62.53** ±3.2*
	3		64.18 ±4.3	**70.26** ±3.7		47.49 ±3.7	**61.77** ±3.6*
	4		64.29 ±4.3	**71.09** ±3.9		47.37 ±3.9	**61.49** ±3.8*
PUMP	1	50.02 ±0.4	81.73 ±6.9	**83.95** ±3.0	49.95 ±0.7	55.72 ±5.4	**77.90** ±4.4*
	2		82.22 ±6.6	**84.55** ±3.2		56.88 ±9.1	**77.71** ±4.2*
	3		82.62 ±6.4	**84.86** ±3.3		55.60 ±10.9	**77.49** ±4.0*
	4		82.89 ±6.3	**85.22** ±3.4		57.09 ±12.5	**77.16** ±3.8*

D.3.2 SEMI-LDP

Table 6: Semi-LDP: detection AUROC (%). Mean ± split SD. Random is shared across budgets.

		Mislabel (25%)			Feature corruption (25%)		
Dataset	α	Random	LDP-KNN	PIVOTS	Random	LDP-KNN	PIVOTS
CODRNA	1	50.08 ±0.5	60.50 ±4.7	**76.17** ±2.9*	50.13 ±0.6	54.26 ±5.0	**69.79** ±2.3*
	2		61.10 ±5.0	**76.01** ±2.8*		52.16 ±6.0	**69.54** ±2.3*
	3		61.56 ±5.0	**75.89** ±2.8*		50.34 ±6.0	**69.35** ±2.3*
	4		62.06 ±5.4	**75.73** ±2.8*		48.63 ±5.1	**69.32** ±2.3*
SVMG1	1	49.85 ±0.6	86.82 ±3.6	**90.13** ±1.0	49.89 ±0.6	63.02 ±3.2	**78.12** ±2.1*
	2		88.98 ±3.2	**90.59** ±1.3		63.47 ±3.6	**77.91** ±1.9*
	3		89.92 ±3.1	**91.00** ±1.4		63.00 ±3.7	**77.71** ±1.7*
	4		90.53 ±3.0	**91.41** ±1.5		61.86 ±3.4	**77.52** ±1.7*
MAGIC	1	50.05 ±1.0	**68.78** ±4.6	66.38 ±2.8	50.06 ±0.7	42.14 ±3.4	**59.99** ±4.8*
	2		**69.73** ±4.6	66.56 ±2.8		40.61 ±3.5	**59.74** ±5.0*
	3		**70.40** ±4.7	66.48 ±2.9		38.45 ±3.5	**59.65** ±5.2*
	4		**71.02** ±4.9*	66.72 ±2.9		37.03 ±3.9	**59.66** ±5.4*
SKIN	1	50.47 ±0.5	96.58 ±0.9	**98.74** ±0.7*	49.88 ±0.5	80.63 ±3.9	**89.95** ±3.6*
	2		98.11 ±0.3	**98.90** ±0.6*		87.84 ±2.2	**89.68** ±3.2
	3		98.68 ±0.3	**98.96** ±0.5		**90.61** ±1.7	89.35 ±3.2
	4		98.97 ±0.2	**99.04** ±0.5		**91.62** ±1.4*	89.75 ±2.5
PHON	1	50.24 ±0.8	73.72 ±3.8	**74.87** ±2.8	49.89 ±0.5	49.31 ±2.8	**61.42** ±3.0*
	2		**75.33** ±2.9	75.24 ±2.6		50.36 ±3.0	**61.42** ±3.0*
	3		**76.57** ±2.3	75.64 ±2.5		51.30 ±3.1	**61.37** ±3.0*
	4		**77.56** ±1.9	75.95 ±2.5		52.45 ±3.2	**61.33** ±3.1*
WATER	1	50.15 ±0.5	48.85 ±1.4	**51.39** ±1.1*	50.08 ±0.6	50.69 ±6.6	**51.92** ±3.1
	2		48.82 ±1.4	**51.39** ±1.1*		49.31 ±6.2	**51.96** ±2.9
	3		48.84 ±1.3	**51.33** ±1.1*		50.10 ±8.0	**52.04** ±3.0
	4		48.69 ±1.2	**51.35** ±1.1*		49.68 ±5.2	**51.90** ±2.9
CALIF	1	49.92 ±0.6	66.55 ±5.9	**66.59** ±4.7	49.82 ±0.7	56.71 ±5.5	**64.78** ±3.9*
	2		**66.90** ±6.1	66.42 ±4.6		57.89 ±7.4	**64.67** ±3.9*
	3		**67.08** ±6.2	66.33 ±4.6		58.44 ±7.3	**64.53** ±3.9
	4		**67.20** ±6.1	66.34 ±4.5		58.65 ±6.8	**64.50** ±3.8*
HOUSE8L	1	49.85 ±0.7	69.20 ±2.4	**72.64** ±1.8*	50.05 ±0.5	51.16 ±7.2	**67.49** ±1.5*
	2		69.36 ±2.4	**72.37** ±1.9		49.33 ±6.1	**67.33** ±1.5*
	3		69.56 ±2.3	**72.11** ±1.6		47.58 ±6.3	**67.35** ±1.5*
	4		69.66 ±2.3	**72.35** ±1.6		45.90 ±7.4	**67.20** ±1.6*
PUMP	1	50.02 ±0.4	**86.60** ±4.3	83.28 ±2.7	49.95 ±0.7	53.14 ±5.3	**78.94** ±2.3*
	2		**86.90** ±4.1	83.28 ±2.8		38.34 ±5.8	**78.57** ±2.7*
	3		**87.17** ±3.9*	83.24 ±2.9		32.19 ±4.2	**78.28** ±3.2*
	4		**87.30** ±3.9*	83.17 ±3.0		28.14 ±4.1	**78.07** ±3.5*

D.3.3 LABEL LDP

Table 7: Label LDP: detection AUROC (%). Mean ± split SD. Random is shared across budgets.

		Mislabel (25%)			Feature corruption (25%)		
Dataset	α	Random	LDP-KNN	PIVOTS	Random	LDP-KNN	PIVOTS
CODRNA	1	50.08 ±0.5	**83.81** ±1.9	82.73 ±2.2	50.13 ±0.6	**69.79** ±1.6*	63.80 ±2.8
	2		**84.59** ±2.0*	82.90 ±2.2		**72.51** ±1.9*	63.57 ±2.5
	3		**84.69** ±2.0*	82.92 ±2.3		**72.93** ±1.8*	63.52 ±2.3
	4		**84.74** ±2.0*	82.94 ±2.3		**73.04** ±1.8*	63.70 ±2.3
SVMG1	1	49.85 ±0.6	**96.30** ±1.3*	94.38 ±1.6	49.89 ±0.6	71.46 ±3.1	**77.51** ±1.5*
	2		**96.59** ±1.2*	94.47 ±1.6		76.52 ±3.1	**77.21** ±1.4
	3		**96.62** ±1.2*	94.50 ±1.6		**78.07** ±3.1	77.14 ±1.5
	4		**96.64** ±1.2*	94.49 ±1.6		**78.56** ±3.1	77.21 ±1.4
MAGIC	1	50.05 ±1.0	**79.66** ±2.4	78.29 ±1.8	50.06 ±0.7	**66.19** ±2.3*	60.24 ±5.0
	2		**80.36** ±2.4*	78.39 ±1.8		**69.03** ±2.5*	60.11 ±5.2
	3		**80.51** ±2.4*	78.42 ±1.8		**69.92** ±2.7*	60.11 ±5.2
	4		**80.54** ±2.5*	78.43 ±1.8		**70.14** ±2.7*	60.12 ±5.2
SKIN	1	50.47 ±0.5	99.19 ±0.7	**99.29** ±0.3	49.88 ±0.5	72.30 ±2.5	**89.97** ±3.0*
	2		**99.36** ±0.7	99.32 ±0.3		78.85 ±2.8	**89.76** ±3.2*
	3		**99.38** ±0.7	99.32 ±0.3		82.26 ±3.1	**89.66** ±3.3*
	4		**99.37** ±0.7	99.33 ±0.3		83.31 ±3.0	**89.64** ±3.2*
PHON	1	50.24 ±0.8	**82.30** ±2.2	81.12 ±2.5	49.89 ±0.5	59.76 ±1.3	**63.27** ±2.3*
	2		**82.91** ±2.1*	81.26 ±2.5		62.99 ±1.6	**63.32** ±2.3
	3		**82.99** ±2.1*	81.31 ±2.5		**63.86** ±1.8	63.29 ±2.2
	4		**83.02** ±2.1*	81.32 ±2.5		**64.06** ±1.9	63.25 ±2.3
WATER	1	50.15 ±0.5	**54.84** ±1.8	52.82 ±1.8	50.08 ±0.6	**53.14** ±1.2*	51.45 ±1.2
	2		**55.05** ±1.8*	52.81 ±1.9		**53.44** ±1.4*	51.62 ±0.9
	3		**55.13** ±1.9*	52.82 ±2.0		**53.35** ±1.4*	51.65 ±0.9
	4		**55.14** ±1.9*	52.81 ±2.0		**53.37** ±1.5*	51.64 ±0.9
CALIF	1	49.92 ±0.6	**77.12** ±5.2*	74.24 ±5.5	49.82 ±0.7	**65.79** ±2.9*	60.73 ±2.9
	2		**77.75** ±5.2*	74.38 ±5.6		**67.82** ±3.7*	60.68 ±3.1
	3		**77.84** ±5.2*	74.41 ±5.6		**68.26** ±3.7*	60.81 ±3.0
	4		**77.85** ±5.2*	74.44 ±5.6		**68.35** ±3.7*	60.93 ±3.1
HOUSE8L	1	49.85 ±0.7	**82.12** ±1.9*	80.38 ±1.9	50.05 ±0.5	**68.71** ±1.7*	65.57 ±1.8
	2		**82.94** ±1.9*	80.50 ±1.9		**71.41** ±1.7*	65.44 ±1.6
	3		**83.03** ±1.9*	80.53 ±1.9		**71.93** ±1.6*	65.36 ±1.5
	4		**83.08** ±1.9*	80.53 ±1.9		**72.05** ±1.6*	65.35 ±1.6
PUMP	1	50.02 ±0.4	90.16 ±2.0	**91.47** ±1.9*	49.95 ±0.7	**77.74** ±1.8*	70.61 ±6.9
	2		90.56 ±2.0	**91.55** ±1.9*		**82.43** ±2.3*	70.39 ±6.8
	3		90.62 ±2.0	**91.57** ±1.9*		**82.83** ±2.3*	70.21 ±6.9
	4		90.64 ±2.0	**91.57** ±1.9*		**82.87** ±2.2*	70.16 ±6.9

D.4 DOWNSTREAM PREDICTION

Random selection and NP-KNN are reference baselines; blank cells repeat their values across budgets. Better recovery or detection need not imply better downstream balanced accuracy.

D.4.1 FULL LDP

Table 8: Full LDP: downstream balanced accuracy (%). Mislabel (25%). Mean ± split SD. Random and NP-KNN are shared across budgets.

Dataset	α	Random	NP-KNN	LDP-KNN	PIVOTS
CODRNA	1	91.58 ±1.6	88.98 ±2.4	63.76 ±8.4	**90.40** ±2.9*
	2			64.02 ±8.8	**90.90** ±2.5*
	3			64.24 ±9.1	**91.21** ±2.5*
	4			64.50 ±9.2	**91.49** ±2.2*
SVMG1	1	89.11 ±1.9	93.53 ±1.8	76.88 ±4.5	**93.62** ±1.6*
	2			78.30 ±4.5	**93.82** ±1.5*
	3			78.88 ±4.5	**93.90** ±1.6*
	4			79.19 ±4.4	**94.07** ±1.7*
MAGIC	1	72.76 ±2.5	75.98 ±2.2	62.89 ±5.7	**69.23** ±3.0
	2			63.15 ±5.7	**70.40** ±3.0*
	3			63.54 ±5.5	**70.85** ±2.7*
	4			63.86 ±5.4	**71.10** ±2.5*
SKIN	1	94.64 ±1.4	92.40 ±4.0	91.99 ±5.6	**94.19** ±1.6
	2			**94.34** ±1.9	94.27 ±1.6
	3			**94.92** ±1.1	94.27 ±1.8
	4			**95.11** ±0.9	94.48 ±1.6
PHON	1	72.97 ±1.8	74.76 ±1.5	72.42 ±2.2	**73.90** ±1.4
	2			73.04 ±1.6	**74.06** ±1.4
	3			73.63 ±1.4	**74.01** ±1.5
	4			73.91 ±1.5	**74.02** ±1.5
WATER	1	50.27 ±1.7	50.30 ±2.7	50.02 ±2.1	**50.80** ±2.3
	2			50.06 ±2.1	**50.81** ±2.4
	3			50.15 ±2.2	**50.76** ±2.5
	4			49.99 ±2.2	**50.59** ±2.5
CALIF	1	80.63 ±1.8	79.97 ±3.4	64.58 ±6.5	**78.26** ±2.3*
	2			64.91 ±6.3	**78.51** ±2.2*
	3			65.02 ±6.1	**78.94** ±2.4*
	4			65.05 ±6.3	**79.22** ±2.2*
HOUSE8L	1	76.64 ±1.2	78.64 ±2.2	64.41 ±5.2	**74.34** ±2.2*
	2			64.09 ±5.4	**74.62** ±2.5*
	3			63.97 ±5.4	**74.89** ±2.5*
	4			63.80 ±5.3	**74.97** ±2.5*
PUMP	1	85.50 ±1.4	86.32 ±1.7	77.96 ±4.7	**83.83** ±2.4*
	2			78.03 ±4.5	**84.10** ±2.1*
	3			78.07 ±4.4	**84.40** ±2.0*
	4			78.24 ±4.4	**84.45** ±2.0*

Table 9: Full LDP: downstream balanced accuracy (%). Feature corruption (25%). Mean ± split SD. Random and NP-KNN are shared across budgets.

Dataset	α	Random	NP-KNN	LDP-KNN	PIVOTS
CODRNA	1	88.69 ±2.5	92.20 ±1.5	70.21 ±8.7	**90.58** ±1.8*
	2			69.94 ±8.7	**90.74** ±1.9*
	3			69.76 ±8.2	**90.78** ±1.9*
	4			68.35 ±8.5	**90.88** ±2.0*
SVMG1	1	92.50 ±1.3	92.80 ±2.5	77.70 ±3.9	**93.60** ±1.2*
	2			78.56 ±4.2	**93.86** ±1.2*
	3			78.79 ±4.5	**93.98** ±1.1*
	4			78.81 ±4.8	**94.12** ±1.1*
MAGIC	1	75.11 ±2.1	75.79 ±2.7	66.64 ±6.4	**72.35** ±2.4*
	2			66.80 ±6.3	**72.62** ±1.9*
	3			67.09 ±6.2	**73.03** ±2.0*
	4			67.45 ±6.0	**73.27** ±1.8*
SKIN	1	93.72 ±1.0	90.76 ±5.0	93.29 ±3.2	**94.14** ±1.8
	2			**94.26** ±1.8	94.14 ±1.7
	3			**94.90** ±1.1	94.07 ±2.2
	4			**95.10** ±1.0	93.93 ±2.2
PHON	1	73.88 ±1.7	74.71 ±1.6	73.55 ±1.5	**74.27** ±1.6
	2			73.67 ±1.3	**74.42** ±1.6
	3			73.82 ±1.3	**74.48** ±1.7
	4			73.96 ±1.4	**74.42** ±1.7
WATER	1	50.50 ±1.1	50.06 ±1.9	50.01 ±1.4	**50.19** ±1.0
	2			50.06 ±1.4	**50.27** ±1.0
	3			50.15 ±1.5	**50.24** ±1.3
	4			50.12 ±1.6	**50.31** ±1.2
CALIF	1	78.31 ±2.5	81.02 ±2.2	69.34 ±5.0	**78.88** ±2.9*
	2			69.40 ±5.5	**78.64** ±2.8*
	3			69.28 ±5.3	**78.48** ±2.9*
	4			69.42 ±5.4	**78.36** ±2.9*
HOUSE8L	1	78.32 ±1.4	78.46 ±2.8	68.65 ±3.8	**77.21** ±1.2*
	2			68.89 ±3.9	**77.21** ±1.3*
	3			68.71 ±3.8	**77.23** ±1.4*
	4			68.64 ±3.7	**77.32** ±1.5*
PUMP	1	85.80 ±1.0	86.38 ±1.4	80.95 ±2.9	**84.90** ±1.9*
	2			81.48 ±3.3	**85.03** ±1.8*
	3			81.59 ±3.3	**85.20** ±1.7*
	4			81.62 ±2.8	**85.25** ±1.5*

D.4.2 SEMI-LDP

Table 10: Semi-LDP: downstream balanced accuracy (%). Mislabel (25%). Mean ± split SD. Random and NP-KNN are shared across budgets.

Dataset	α	Random	NP-KNN	LDP-KNN	PIVOTS
CODRNA	1	91.58 ±1.6	88.98 ±2.4	64.77 ±10.6	**89.17** ±3.3*
	2			66.13 ±10.8	**89.21** ±3.1*
	3			66.33 ±10.4	**89.13** ±3.3*
	4			66.66 ±10.3	**89.13** ±3.3*
SVMG1	1	89.11 ±1.9	93.53 ±1.8	77.98 ±4.6	**93.44** ±1.2*
	2			79.23 ±4.5	**93.50** ±1.2*
	3			80.04 ±4.3	**93.59** ±1.1*
	4			80.87 ±4.1	**93.67** ±1.1*
MAGIC	1	72.76 ±2.5	75.98 ±2.2	67.88 ±5.2	**67.93** ±3.5
	2			**68.41** ±5.1	68.07 ±3.3
	3			**68.76** ±4.8	67.91 ±3.1
	4			**69.04** ±5.0	68.37 ±3.1
SKIN	1	94.64 ±1.4	92.40 ±4.0	93.29 ±2.8	**94.77** ±1.8*
	2			**94.81** ±1.0	94.62 ±1.9
	3			**94.99** ±1.0	94.84 ±1.9
	4			**95.08** ±1.0	94.65 ±2.0
PHON	1	72.97 ±1.8	74.76 ±1.5	71.06 ±3.8	**73.27** ±1.7*
	2			72.04 ±2.2	**73.42** ±1.6
	3			72.55 ±1.6	**73.55** ±1.5*
	4			72.98 ±1.5	**73.67** ±1.6
WATER	1	50.27 ±1.7	50.30 ±2.7	49.96 ±2.0	**50.89** ±1.2
	2			49.95 ±2.4	**50.96** ±1.0
	3			49.83 ±2.4	**50.92** ±1.1
	4			49.98 ±2.5	**50.87** ±1.3
CALIF	1	80.63 ±1.8	79.97 ±3.4	66.64 ±5.3	**75.56** ±4.1*
	2			66.55 ±5.6	**75.51** ±4.1*
	3			66.52 ±5.8	**75.47** ±4.0*
	4			66.60 ±5.8	**75.61** ±4.0*
HOUSE8L	1	76.64 ±1.2	78.64 ±2.2	66.44 ±3.9	**75.17** ±1.9*
	2			66.43 ±3.7	**74.99** ±2.0*
	3			66.38 ±3.5	**75.05** ±1.9*
	4			66.55 ±3.5	**75.52** ±1.4*
PUMP	1	85.50 ±1.4	86.32 ±1.7	80.20 ±4.1	**80.99** ±2.4
	2			80.82 ±3.6	**80.97** ±2.4
	3			**81.17** ±3.3	80.85 ±2.6
	4			**81.41** ±3.1	80.76 ±2.8

Table 11: Semi-LDP: downstream balanced accuracy (%). Feature corruption (25%). Mean ± split SD. Random and NP-KNN are shared across budgets.

Dataset	α	Random	NP-KNN	LDP-KNN	PIVOTS
CODRNA	1	88.69 ±2.5	92.20 ±1.5	74.73 ±9.0	**91.20** ±2.0*
	2			73.49 ±8.3	**91.12** ±1.8*
	3			73.39 ±9.2	**90.92** ±1.9*
	4			72.86 ±9.0	**90.93** ±1.8*
SVMG1	1	92.50 ±1.3	92.80 ±2.5	77.32 ±4.8	**93.48** ±1.4*
	2			78.98 ±5.0	**93.38** ±1.5*
	3			79.99 ±4.5	**93.50** ±1.6*
	4			81.40 ±4.1	**93.65** ±1.4*
MAGIC	1	75.11 ±2.1	75.79 ±2.7	70.84 ±4.4	**72.26** ±2.1
	2			71.15 ±4.0	**71.87** ±2.6
	3			**71.65** ±3.8	71.54 ±2.9
	4			**71.90** ±3.3	71.88 ±2.8
SKIN	1	93.72 ±1.0	90.76 ±5.0	91.86 ±4.7	**94.23** ±1.8
	2			**94.75** ±1.2	94.20 ±2.0
	3			**94.98** ±0.9	94.11 ±2.0
	4			**94.87** ±1.2	94.15 ±2.3
PHON	1	73.88 ±1.7	74.71 ±1.6	72.56 ±1.4	**74.25** ±1.8*
	2			72.61 ±1.6	**74.30** ±1.7*
	3			72.78 ±1.8	**74.36** ±1.6*
	4			72.78 ±1.8	**74.40** ±1.6*
WATER	1	50.50 ±1.1	50.06 ±1.9	50.31 ±1.5	**50.52** ±1.1
	2			50.28 ±1.5	**50.44** ±1.0
	3			50.22 ±1.4	**50.57** ±1.0
	4			50.24 ±1.6	**50.53** ±1.0
CALIF	1	78.31 ±2.5	81.02 ±2.2	70.73 ±5.3	**78.45** ±3.4*
	2			70.87 ±5.1	**78.59** ±3.2*
	3			71.05 ±5.0	**78.64** ±3.3*
	4			71.03 ±5.1	**78.63** ±3.3*
HOUSE8L	1	78.32 ±1.4	78.46 ±2.8	69.95 ±3.2	**77.72** ±1.4*
	2			69.99 ±3.1	**77.60** ±1.5*
	3			70.17 ±3.4	**77.66** ±1.5*
	4			70.18 ±3.2	**77.73** ±1.5*
PUMP	1	85.80 ±1.0	86.38 ±1.4	81.88 ±2.6	**83.44** ±2.2
	2			82.27 ±2.6	**83.47** ±2.2
	3			82.20 ±2.6	**83.43** ±2.2
	4			82.41 ±2.3	**83.35** ±2.1

D.4.3 LABEL LDP

Table 12: Label LDP: downstream balanced accuracy (%). Mislabel (25%). Mean ± split SD. Random and NP-KNN are shared across budgets.

Dataset	α	Random	NP-KNN	LDP-KNN	PIVOTS
CODRNA	1	91.58 ±1.6	88.98 ±2.4	88.99 ±2.3	**91.29** ±2.8
	2			88.84 ±2.5	**91.31** ±2.9
	3			88.91 ±2.6	**91.29** ±2.8
	4			88.92 ±2.6	**91.31** ±2.8
SVMG1	1	89.11 ±1.9	93.53 ±1.8	93.59 ±1.7	**93.60** ±0.9
	2			**93.63** ±1.8	93.60 ±0.8
	3			93.56 ±1.8	**93.61** ±0.9
	4			93.53 ±1.8	**93.60** ±0.8
MAGIC	1	72.76 ±2.5	75.98 ±2.2	**75.77** ±2.2*	73.97 ±2.1
	2			**75.83** ±2.2*	74.02 ±2.0
	3			**75.85** ±2.3*	74.09 ±2.1
	4			**75.89** ±2.2*	74.05 ±2.0
SKIN	1	94.64 ±1.4	92.40 ±4.0	92.92 ±3.4	**94.07** ±2.2
	2			92.74 ±3.7	**93.64** ±2.9
	3			92.56 ±3.8	**93.49** ±3.2
	4			92.48 ±3.9	**93.44** ±3.4
PHON	1	72.97 ±1.8	74.76 ±1.5	**74.61** ±1.4	74.56 ±2.0
	2			**74.68** ±1.6	74.52 ±2.0
	3			**74.72** ±1.5	74.54 ±2.0
	4			**74.75** ±1.5	74.45 ±2.0
WATER	1	50.27 ±1.7	50.30 ±2.7	50.26 ±2.3	**50.47** ±2.2
	2			50.28 ±2.4	**50.52** ±2.4
	3			50.27 ±2.4	**50.51** ±2.4
	4			50.23 ±2.5	**50.55** ±2.5
CALIF	1	80.63 ±1.8	79.97 ±3.4	**79.75** ±3.2	77.70 ±3.4
	2			**79.94** ±3.2	77.68 ±3.3
	3			**79.94** ±3.3	77.72 ±3.3
	4			**79.98** ±3.3	77.62 ±3.3
HOUSE8L	1	76.64 ±1.2	78.64 ±2.2	**78.72** ±2.1	77.73 ±2.0
	2			**78.69** ±2.2	77.73 ±2.2
	3			**78.74** ±2.3	77.73 ±2.2
	4			**78.68** ±2.2	77.75 ±2.1
PUMP	1	85.50 ±1.4	86.32 ±1.7	**86.06** ±1.8	85.26 ±1.7
	2			**86.19** ±1.8*	85.32 ±1.6
	3			**86.30** ±1.8	85.35 ±1.6
	4			**86.34** ±1.7*	85.37 ±1.6

Table 13: Label LDP: downstream balanced accuracy (%). Feature corruption (25%). Mean ± split SD. Random and NP-KNN are shared across budgets.

Dataset	α	Random	NP-KNN	LDP-KNN	PIVOTS
CODRNA	1	88.69 ±2.5	92.20 ±1.5	**92.13** ±1.9	90.52 ±2.1
	2			**92.33** ±1.7	90.63 ±2.5
	3			**92.36** ±1.6	90.62 ±2.8
	4			**92.27** ±1.5	90.63 ±2.7
SVMG1	1	92.50 ±1.3	92.80 ±2.5	93.10 ±1.7	**93.86** ±1.4
	2			92.97 ±2.2	**94.00** ±1.4
	3			92.84 ±2.4	**94.08** ±1.4
	4			92.88 ±2.4	**94.05** ±1.4
MAGIC	1	75.11 ±2.1	75.79 ±2.7	**75.91** ±2.7	74.31 ±2.3
	2			**75.91** ±2.6	74.32 ±2.3
	3			**75.83** ±2.6	74.33 ±2.3
	4			**75.88** ±2.6	74.34 ±2.3
SKIN	1	93.72 ±1.0	90.76 ±5.0	91.71 ±4.0	**93.71** ±2.9*
	2			90.99 ±4.5	**93.89** ±2.9*
	3			90.68 ±4.9	**93.87** ±3.0*
	4			90.44 ±5.0	**93.95** ±3.0*
PHON	1	73.88 ±1.7	74.71 ±1.6	74.39 ±1.5	**75.02** ±2.1
	2			74.66 ±1.4	**75.11** ±2.1
	3			74.58 ±1.4	**75.13** ±2.1
	4			74.67 ±1.4	**75.07** ±2.1
WATER	1	50.50 ±1.1	50.06 ±1.9	50.12 ±2.0	**50.53** ±2.4
	2			50.01 ±2.2	**50.66** ±2.4
	3			50.00 ±2.3	**50.60** ±2.7
	4			50.03 ±2.3	**50.58** ±2.7
CALIF	1	78.31 ±2.5	81.02 ±2.2	**80.67** ±2.4*	77.91 ±3.5
	2			**80.70** ±2.2	77.94 ±3.7
	3			**80.73** ±2.1	77.95 ±3.6
	4			**80.79** ±2.1	77.90 ±3.7
HOUSE8L	1	78.32 ±1.4	78.46 ±2.8	**78.75** ±2.2	77.27 ±1.5
	2			**78.61** ±2.5	77.21 ±1.7
	3			**78.47** ±2.6	77.02 ±1.9
	4			**78.50** ±2.7	77.10 ±1.8
PUMP	1	85.80 ±1.0	86.38 ±1.4	**86.27** ±1.3*	85.51 ±1.5
	2			**86.25** ±1.5*	85.40 ±1.3
	3			**86.31** ±1.4*	85.37 ±1.2
	4			**86.26** ±1.5*	85.40 ±1.3

E Additional Simulation Results

E.1 Shared setup

The population model and 2D settings are specified in Section 5.1. The 1D study uses full LDP with the same α and N grids, $M = 1024N$, and $\tau_N = N^{-1/4}$. We average squared errors across sampled providers and use 2000 outer repetitions in 1D and 200 in 2D. Non-private targets use the original feature distances at each τ_N. The 2D semi-LDP estimator bins only private features; the 2D label-LDP estimator uses randomized response and exact feature neighborhoods.

E.2 Risk and uncertainty

For $a_i = \mathbb{E}_Q \widetilde{\phi}_i$ and $b_i = \mathbb{E}_Q \widetilde{\phi}_i^2$, conditional risk is

$$\mathbb{E}[(\widetilde{\Phi}_i - \Phi_i^\star)^2 \mid D, W] = (a_i - \Phi_i^\star)^2 + (b_i - a_i^2)/M. \tag{99}$$

Write $B_i = (a_i - \Phi_i^\star)^2$ and $V_i = b_i - a_i^2$; validation-size curves use their outer averages $\bar{B} + \bar{V}/M$. Integrals are analytic in 1D and estimated with independent scrambled Sobol samples in 2D. The N-curves show mean ± 1 outer SE; slope intervals use a paired outer bootstrap.

E.3 Validation-size plateau

At $N = 32768$, we vary M on a geometric grid and plot $\bar{B} + \bar{V}/M$. The full-LDP panel in Figure 2 and semi-LDP curves in Figure 3 approach a non-validation error floor; some label-LDP curves remain validation-limited over the plotted range.

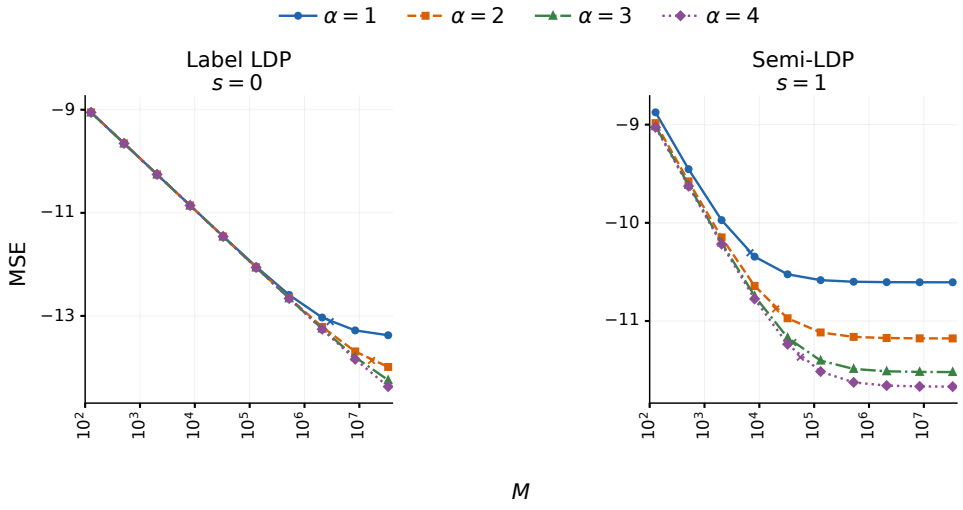

Figure 3: **Validation-size scaling in 2D under label LDP and semi-LDP.** Mean $B + V/M$ at $N = 32768$ for $s = 0, 1$. Crosses mark $M = \overline{V}/\overline{B}$. Vertical tick labels are base-10 exponents; panels have separate MSE scales.

E.4 One-dimensional results

The full-LDP slopes in Figure 4 are near -2.1 across privacy budgets. The selected-bandwidth rate has polynomial exponent $\gamma(d, s) = 2 + \frac{2}{s+2}(1 - \frac{d+s}{4d})$; finite-range slopes also reflect logarithms, grid rounding, and validation error.

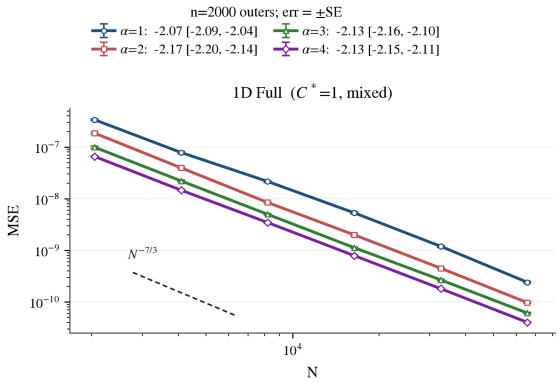

Figure 4: **1D Full LDP: provider-sample scaling.** The shared setup applies. Legends give slopes and 95% bootstrap intervals.

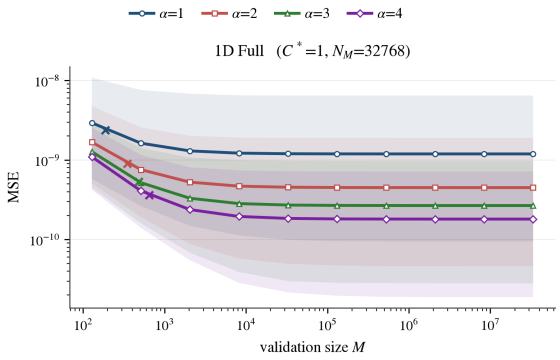

Figure 5: **1D Full LDP: validation-size plateau.** Bands show outer-repetition ranges, not confidence intervals for the mean.

E.5 Controlled count ablation

Direct plug-in, joint fitting, and PIVOTS use the same 1D full-LDP records, reports, and targets at $N \in \{2048, 8192, 32768\}$ and $\alpha \in \{1, 3, 4\}$. Direct plug-in clips and rounds raw counts; joint fitting enforces feasible integer label counts; PIVOTS additionally stabilizes the fitted total. A false zero has a positive non-private binned count but a zero estimated count. At $N = 8192$ and $\alpha = 1$, its observed query-mass frequency is 6.82%, 2.33%, and 0% for the three methods, respectively. PIVOTS has the lowest mean MSE in Figures 6 and 7; the gap narrows as α increases.

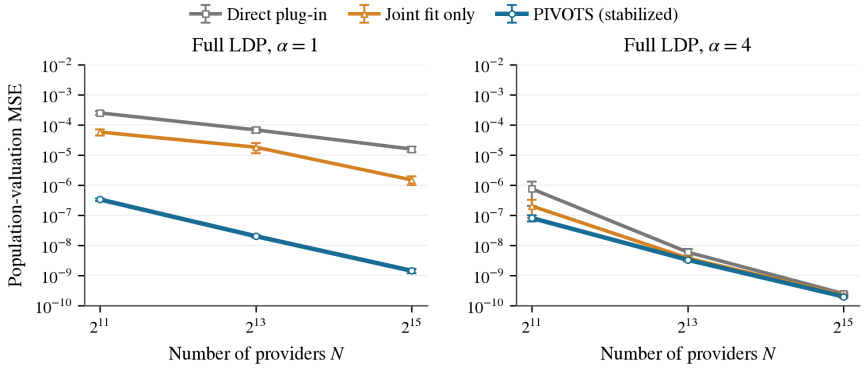

Figure 6: **Stabilization ablation under 1D Full LDP.** Mean MSE ± 1 outer SE over 200 repetitions. All methods share records, reports, and targets.

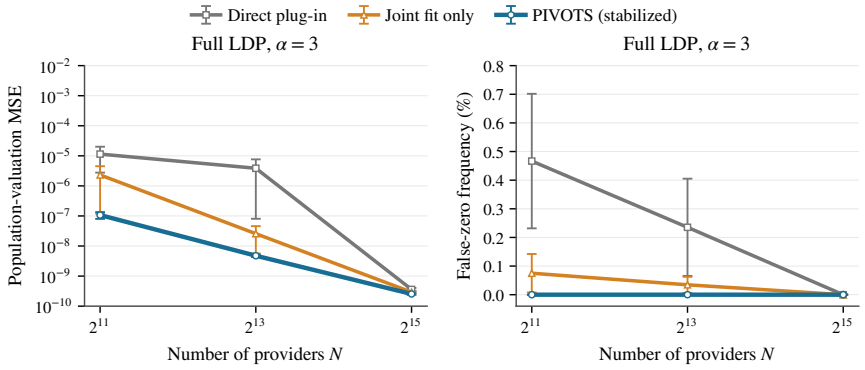

Figure 7: **Count ablation at $\alpha = 3$.** Identical reports across methods, with ± 1 outer SE.

E.6 SCALING WITHOUT STABILIZATION

We remove only anchor mixing from the joint-fit estimator and compare paired runs. Mean MSE rises in all 72 feature-private settings; label-LDP results coincide because their neighborhood counts are exact. Table 14 reports finite-sample slopes; these comparisons do not establish an asymptotic advantage.

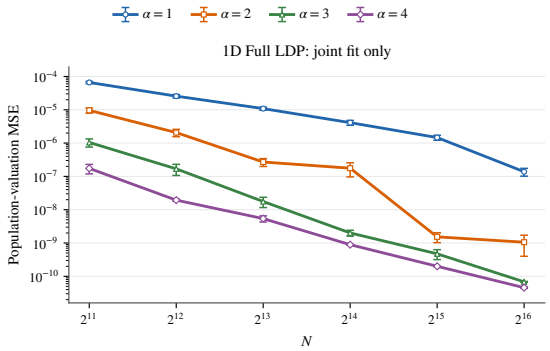

Figure 8: **1D Full LDP without stabilization.** Compare with Figure 4. All ablation curves use four budgets and ± 1 outer SE.

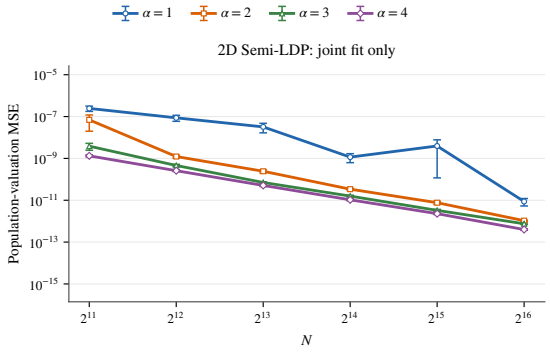

Figure 9: **2D Semi-LDP without stabilization.** Compare with Figure 2.

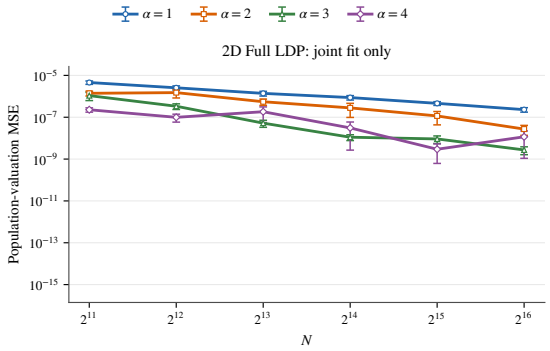

Figure 10: **2D Full LDP without stabilization.** Compare with Figure 2.

Table 14: **N-rate slopes with and without anchor mixing.** Slopes regress log mean population-valuation MSE on log N over the six original sample sizes. Brackets are paired outer-bootstrap 95% percentile intervals (5000 resamples). The last column gives the range of joint-only/PIVOTS mean-MSE ratios across these sizes. These are finite-range descriptive slopes for a changing-τ_N target, not tests of a minimax exponent.

Setting	α	Without anchor mixing	PIVOTS	MSE ratio range
1D Full	1	-1.67 [-1.79, -1.55]	-2.07 [-2.09, -2.04]	196–1.23e+03
1D Full	2	-2.79 [-3.28, -2.56]	-2.17 [-2.20, -2.15]	3.4–88.6
1D Full	3	-2.80 [-2.95, -2.62]	-2.13 [-2.16, -2.10]	1.12–10.6
1D Full	4	-2.34 [-2.45, -2.20]	-2.13 [-2.15, -2.11]	1.11–2.65
2D $s = 0$	1	-2.51 [-2.57, -2.45]	-2.51 [-2.57, -2.45]	1–1
2D $s = 0$	2	-2.54 [-2.58, -2.51]	-2.54 [-2.58, -2.51]	1–1
2D $s = 0$	3	-2.59 [-2.62, -2.57]	-2.59 [-2.62, -2.57]	1–1
2D $s = 0$	4	-2.59 [-2.60, -2.58]	-2.59 [-2.60, -2.58]	1–1
2D $s = 1$	1	-2.63 [-3.27, -2.39]	-2.18 [-2.26, -2.10]	2.48–158
2D $s = 1$	2	-3.00 [-3.21, -2.59]	-2.43 [-2.52, -2.32]	1.08–14.1
2D $s = 1$	3	-2.43 [-2.57, -2.26]	-2.29 [-2.37, -2.20]	1.06–2.01
2D $s = 1$	4	-2.32 [-2.39, -2.25]	-2.29 [-2.36, -2.23]	1.05–1.18
2D $s = 2$	1	-0.85 [-1.00, -0.69]	-1.90 [-1.97, -1.83]	302–1.25e+04
2D $s = 2$	2	-1.16 [-1.53, -0.89]	-1.91 [-1.99, -1.83]	158–2.38e+03
2D $s = 2$	3	-1.74 [-2.04, -1.45]	-1.94 [-2.03, -1.86]	74.7–273
2D $s = 2$	4	-1.12 [-2.19, -0.80]	-1.90 [-2.02, -1.78]	31–1.03e+03

F Extended Related Work

F.1 Data Valuation and Tractable Shapley Computation

Ghorbani & Zou (2019) develop Data Shapley for supervised learning, and Jia et al. (2019b) study computational approximations to Shapley-based data valuation. Kwon & Zou (2022) propose Beta Shapley by changing the weights assigned to subset sizes, Wang & Jia (2023a) analyze the robustness of Data Banzhaf to noisy utility evaluations, and Kwon & Zou (2023) estimate data values from out-of-bag evaluations of bagging models.

Jia et al. (2019a) compute Shapley values for nearest-neighbor utilities using distance ordering and recursion. Wang & Jia (2023b) study alternative KNN utility definitions and derive corresponding Shapley formulas, while Wang et al. (2023) introduce TKNN-Shapley and its local-count closed form. Wang et al. (2025) attribute data contributions along a model's training trajectory. Ghorbani et al. (2020) define Distributional Shapley by averaging a record's value over sampled training datasets.

F.2 Privacy-Preserving Valuation and Pre-Acquisition Appraisal

Wang et al. (2023) show that KNN-Shapley scores can reveal information about other records and develop DP-TKNN-Shapley by perturbing counts held by a trusted curator. Xu et al. (2022) combine secure computation with forward influence functions to appraise candidate datasets before acquisition. Tian et al. (2022) use a learned utility predictor and multiparty computation for private Shapley-based valuation and payment.

F.3 Locally Private Statistical Estimation

Duchi et al. (2018) develop information-theoretic lower-bound techniques for local privacy and characterize optimal rates for several estimation problems. Kairouz et al. (2016b) study the design of local randomizers for information-theoretic utility criteria. Wang et al. (2017) develop frequency-estimation protocols based on optimized unary encoding and local hashing, and Acharya et al. (2019) introduce Hadamard Response for distribution estimation with low communication cost.

Berrett et al. (2021) construct universally consistent regression and classification procedures from privatized partition statistics. Györfi & Kroll (2025) analyze a partition-based private regression estimator and establish matching rates under their nonparametric assumptions.

F.4 Public Features, Partial Privacy, and Label Privacy

Curmei et al. (2023) use public item features in private matrix factorization. Krichene et al. (2024) study private learning with joined public and private features, including linear regression. Chua et al. (2024) study ad prediction with semi-sensitive features. Shen et al. (2023) study classification with a designated private feature block and public labels. Ma et al. (2026) analyze nonparametric regression under semi-LDP with aligned and personalized public-feature splits.

Chaudhuri & Hsu (2011) study sample-complexity distinctions associated with label privacy. Ghazi et al. (2021) develop randomized response using label priors for private deep learning, and Malek Esmaeili et al. (2021) study Laplace-based Bayesian inference and a PATE-based approach. Ghazi et al. (2023) use randomized response on bins for regression, while Badanidiyuru Varadaraja et al. (2023) develop unbiased label randomizers. Ghazi et al. (2024) use public-feature projections in a central-DP training procedure, and Zhao et al. (2025b) study vector approximation under label privacy.

Wang & Xu (2019) contrast sparse regression with private features and responses against regression with only private responses. Xu et al. (2023) study binary classification under local label privacy using randomized response. Zhao et al. (2025a) establish theoretical limits for label-private nonparametric classification and regression in both local and central models. Busa-Fekete et al. (2023) examine label privacy in private training-data release and the role of feature–label dependence.